

**SEARCH REQUEST FORM**

Scientific and Technical Information Center

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
 Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
 Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

**If more than one search is submitted, please prioritize searches in order of need.**

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

\*\*\*\*\*  
**STAFF USE ONLY**

Searcher: \_\_\_\_\_ **Contact:** Shoppard  
 Searcher Phone #: rel. 308-4499  
 Searcher Location: \_\_\_\_\_  
 Date Searcher Picked Up: \_\_\_\_\_  
 Date Completed: 9/3/00  
 Searcher Prep & Review Time: \_\_\_\_\_  
 Clerical Prep Time: \_\_\_\_\_  
 Online Time: \_\_\_\_\_

**Type of Search**

NA Sequence (#) \_\_\_\_\_  
 AA Sequence (#) \_\_\_\_\_  
 Structure (#) \_\_\_\_\_  
 Bibliographic \_\_\_\_\_  
 Litigation \_\_\_\_\_  
 Fulltext \_\_\_\_\_  
 Patent Family \_\_\_\_\_  
 Other \_\_\_\_\_

**Vendors and cost where applicable**

STN \_\_\_\_\_  
 Dialog \_\_\_\_\_  
 Questel/Orbit \_\_\_\_\_  
 Dr.Link \_\_\_\_\_  
 Lexis/Nexis \_\_\_\_\_  
 Sequence Systems \_\_\_\_\_  
 WWW/Internet \_\_\_\_\_  
 Other (specify) \_\_\_\_\_



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2000, 23:12:56 ; Search time 2111.72 Seconds  
(without alignments)  
2743.978 Million cell updates/sec

Title: US-09-464-528-6  
Perfect score: 1314  
Sequence: 1 tctagatcaaacacacatcc.....ttttgaagtataaccatgg 1314

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*

44: gb\_est25: \*  
45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: em\_est20: \*  
53: em\_est21: \*  
54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*  
59: gb\_est33: \*  
60: gb\_est34: \*  
61: gb\_est35: \*  
62: gb\_est36: \*  
63: gb\_est37: \*  
64: gb\_est38: \*  
65: em\_est27: \*  
66: em\_est28: \*  
67: em\_est29: \*  
68: em\_est30: \*  
69: gb\_est39: \*  
70: gb\_est40: \*  
71: gb\_est41: \*  
72: gb\_est42: \*  
73: gb\_est43: \*  
74: gb\_est44: \*  
75: em\_est31: \*  
76: em\_est32: \*  
77: em\_est33: \*  
78: em\_est34: \*  
79: gb\_est45: \*  
80: gb\_est46: \*  
81: gb\_est47: \*  
82: em\_est35: \*  
83: em\_est36: \*  
84: em\_est37: \*  
85: gb\_est48: \*  
86: gb\_est49: \*  
87: gb\_est50: \*  
88: gb\_est51: \*  
89: gb\_est52: \*  
90: gb\_est53: \*  
91: gb\_est54: \*  
92: gb\_est55: \*  
93: gb\_gss1: \*  
94: gb\_gss2: \*  
95: gb\_gss3: \*  
96: gb\_gss4: \*  
97: em\_gss1: \*  
98: em\_gss2: \*  
99: em\_gss3: \*  
100: em\_gss4: \*  
101: gb\_gss5: \*  
102: gb\_gss6: \*  
103: gb\_gss7: \*  
104: gb\_gss8: \*  
105: gb\_gss9: \*  
106: em\_gss5: \*  
107: em\_gss6: \*  
108: em\_gss7: \*  
109: em\_gss8: \*  
110: em\_gss9: \*  
111: em\_gss10: \*  
112: em\_gss11: \*  
113: gb\_gss10: \*  
114: gb\_gss11: \*  
115: em\_gss12: \*  
116: gb\_gss12: \*

```

source
1. 843
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR26H19"
/note="end : TET3"
131 a 102 c 259 g 141 t 210 others
BASE COUNT
ORIGIN

```

Query Match 6.4%; Score 84.6; DB 122; Length 843;  
Best Local Similarity 39.2%; Pred. No. 1.1e-05;  
Matches 121; Conservative 65; Mismatches 123; Indels 0; Gaps 0;

Qy	1	tctagatcaaacactcacacccaaacataacaatcgggatatctccttaccaatcatcataatta	60
Db	831	TAWAWATAAATATATATWTWTTTTTATAATTDDMTWMDAAATAMWTGTATWATTTWTITA	772
Qy	61	ttttgggtcaaatatttaattcatttatctttttaagaatatcaattaaagaanaattaaaagattttt	120
Db	771	TTTTTAWTWTTTTTTTWTFATTATTAAATTTWTATATWTWTATWTATATTTWAATTTWMTWTAT	712
Qy	121	taaaaaaatgcataaaaattattatttcacgatcttttcacacatttgatttttgataataa	180
Db	711	ATTAWATATATATAWMAATAAAATTAATTTATTTWNAWWWMTTATTTTATTTTATWTWTWATAW	652
Qy	181	atatatttttttaattcttcaaaaagtgtgcgaagacaccttatagacacagctcttgtt	240
Db	651	TTWTWTTWATAWAWATTTAWAATAWATATAWAAATTTAAWTTTTWTRIAAAAAANWWNAANA	592
Qy	241	ctgtttcaaaaagcattcattcaattcaacattaaaaaatttaaactactaacagtagaa	300
Db	591	AATWATATAAAAANWWNATAMAHTTTTATMRAATWNAWATAATAAHWAAMAATAAANWNA	532
Qy	301	tctctcttgt	309
Db	531	ATWWTATWAT	523

RESULT 3  
CNS016LI/c  
LOCUS  
DEFINITION  
CNS016LI 1101 bp DNA GSS 26-JUL-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN16D22 of DrosBAC library from Drosophila melanogaster (fru  
fly), genomic survey sequence.  
ALI06896  
ALI06896.1 GI:5624374  
GSS.  
fruit fly.  
Drosophila melanogaster  
ORGANISM  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequenc  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cn  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part  
collaboration with the European Drosophila genome project (EDG  
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MR  
project grant. The DNA was prepared from embryos by Alain Buch  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.  
Location/Qualifiers  
1. .1101  
/organism="Drosophila melanogaster"  
/plasmid="pBelobAC11"  
/db\_xref="taxon:7227"  
/clone\_lib="DrosBAC"

```
/clone_lib="DrosBAC"
```





```

Matches 153; Conservative 58; Mismatches 158; Indels 2; Gaps 1;
Qy 35 tatctcctaccatacctaataatttttgggttaataataatacattatttttaagat 94
Db 548 TCCTTTTWTATATATATATATATATATATATATATATATATATATATATATAT 607
Qy 95 attaatgaagaatttaaaagtatttttaaaagtataaaattatatttcaatgatt 154
Db 608 AAWAATATATATATATATATATATATATATATATATATATATATATATATAT 667
Qy 155 ttctacatttgatttttgataataataataatttttttttttttttttttttttt 214
Db 668 TTTTWTATATATATATATATATATATATATATATATATATATATATATATAT 727
Qy 215 agacattattagacatagcttcttctgtttacaaaagcattcattcaatacatta 274
Db 728 ATTATATATATATATATATATATATATATATATATATATATATATATATAT 787
Qy 275 aaaaatttt--aatactaacagtagaactcttctgtgagtggtggtgagtaggcaacc 332
Db 788 AAWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 847
Qy 333 tggcattgaacagagagagagagagagagagagagagagagagagagagagagag 392
Db 848 TAWAWAWAWAWAWAWAWAWAWAWAWAWAWAWAWAWAWAWAWAWAWAWAWAWAW 907
Qy 393 acaatacaaaa 403
Db 908 NCWAATHAWAA 918

RESULT 6
LOCUS CNS016CO 1200 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN15E04 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106578
VERSION AL106578.1 GI:5622626
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1200)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
source
1..1200
/organism="Drosophila melanogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN15E04"
/note="end : T7"
BASE COUNT 351 a 188 c 167 g 307 t 187 others
ORIGIN

Query Match 6.1%; Score 79.8; DB 123; Length 1200;
Best Local Similarity 42.8%; Pred. No. 7.3e-05;

```

```

Matches 115; Conservative 57; Mismatches 94; Indels 3; Gaps 1;
Qy 48 atcatacctaatttttgggttaataataatacattatttttaagataataaagaaa 107
Db 1197 ATTWATTTATATATATATATATATATATATATATATATATATATATATATAT 1138
Qy 108 ttaaaagatttttttaaaagtataaaattatatttcttcaatgatttttccatattg 167
Db 1137 ATWTATATATATATATATATATATATATATATATATATATATATATATAT 1078
Qy 168 atttgataataataataatttttttttttttttttttttttttttttttttttat 224
Db 1077 TWTTTTWTATATATATATATATATATATATATATATATATATATATATATAT 1018
Qy 225 tagacatagcttcttctgtttacaaaagcattcattcaatacaataaataatttt 284
Db 1017 TTTTATATATATATATATATATATATATATATATATATATATATATATAT 958
Qy 285 aatactaacagtagaactcttctgtgagt 313
Db 957 AATTWAAAAAAWATTTTWTTRTTWWAAT 929

RESULT 7
LOCUS CNS00DKY 928 bp DNA GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL071865
VERSION AL071865.1 GI:4948170
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 928)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Mammosier in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
ECORI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Location/Qualifiers
1..928
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR27A24"
/note="end : T7"
BASE COUNT 262 a 70 c 84 g 321 t 191 others
ORIGIN

Query Match 6.0%; Score 78.8; DB 122; Length 928;
Best Local Similarity 29.7%; Pred. No. 0.00011;
Matches 91; Conservative 99; Mismatches 116; Indels 0; Gaps 0;

```





[illegible]

[illegible]

RESULT 14  
CNS000DKY 928 bp DNA GSS 04-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #  
DEFINITION BACR27A24 of RCI-98 library from *Drosophila melanogaster* (fruit fly); genomic survey sequence.  
AL071865  
AL071865.1 GI:4948170  
GSS.  
KEYWORDS  
SOURCE fruit fly.  
ORGANISM *Drosophila melanogaster*  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 928)  
REFERENCE

BASE COUNT	262 a	70 c	84 g	321 t	191 others
ORIGIN					

		Query Match	5.7%	Score 74.6;	DB 122;	Length 928;
		Best Local Similarity	32.8%;	Pred. No.	0.00059;	
		Matches	96;	Conservative	91;	Mismatches 103; Indels 3; Gaps 1;
Qy	35	tatctccttccaacatactaactatttttggggttaaatattaatcaattattttttaagat	94			
Db	557	TWTWTTTTTTTWWTTTAAATTTATTATTATTAATAAAGGTAAAGGTTTATATATWTTAAATA	616			
Qy	95	attaattaagaagaattaaaaggatttttttcacaaaaatgataaaatatcatattcatgat	154			
Db	617	WMAAATATWMTATATAATATWTTTWTTWTTWTTATATATAWAANAANAA---AAAAAW	673			
Qy	155	ttccatacatattgatttgataacaatacataatttttttaaattttcttcaaaaaagtgcga	214			
Db	674	AATAAAAAAATTTTAAWAAWTAATAAAAANAAATTTATWTTTTTTTTTTTWTATWTTAAWATA	733			
Qy	215	agacacctatttagacatatgtcttgtctgtttaccaaaagcattcatcatttaaacatta	274			
Db	734	WAAATATWMMWMTWTDGNKNNNNAWWWMMMMMWWMAAAAMWWWMMMMMWMAAAA	793			
Qy	275	aaaaatttatctaactaacagtagaatactctctcgagctggcgtaggtagg	327			
Db	794	AAAAAAAAAAWAAAAAADWDODDDDWKAARKKKKKKKKKKKKKKKKKKKKKKKKK	846			

RESULT 15  
CNS00FUH/c  
LOCUS  
DEFINITION  
                  996 bp      DNA                  03-JUN-1999  
                  GSS  
Drosophila melanogaster genome survey sequence TET3 end of BAC:  
BACR31021 of RC1-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION  
VERSION          AL071063.1  GI:4951105  
KEYWORDS  
SOURCE           fruit fly.  
                  Drosophila melanogaster  
ORGANISM  
                  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
                  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
                  Muscumorphia; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
                  1 (bases 1 to 996)  
                  Genoscope.  
AUTHORS  
                  Direct Submission  
TITLE  
                  Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
JOURNAL          BP 191 91006 EVRY cedex - FRANCE (E-mail : segr@genoscope.cns.fr  
                  - Web : www.genoscope.cns.fr)

```

BASE COUNT      383 a      164 c      81 g      171 t      197 others
ORIGIN

Query Match      5.7%      Score 74.4;      DB 122;      Length 996;
Best Local Similarity 41.0%;      Pred. No. 0.00064;

```

```
Matches 109; Conservative 50; Mismatches 107; Indels 0; Gaps 0;
Qy 26 taacatggatctctccatccatccatctatttttgggttaaatatcattatcattat 85
   | : | : | | : | | | | | | | | | | | | | | | | | | | | |
Db 863 TTAAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 804
   | : | : | | : | | | | | | | | | | | | | | | | | | | | |
Qy 86 ttttaagatataataaagaatttaaaagattttttaaaaaaataataatata 145
   | : | : | | : | | | | | | | | | | | | | | | | | | | | |
Db 803 AAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 744
   | : | : | | : | | | | | | | | | | | | | | | | | | | | |
Qy 146 ttcatgatttttccatcacatttgatttgataataataatatttttttcttcttaaaa 205
   | | : | | | | | : | : | | | | | | | | | | | | | | | | |
Db 743 TTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 684
   | | : | | | | | : | : | | | | | | | | | | | | | | | | |
Qy 206 aatgttgcaagacacttattgacatagctctgtctgtttcacaaaaagcattcattt 265
   | | : | | | | | : | : | | | | | | | | | | | | | | | | |
Db 683 TTTATTTAAATGATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 624
   | | : | | | | | : | : | | | | | | | | | | | | | | | | |
Qy 266 aatacattaaaaaataatttaacta 291
   | | : | | : | | | | | | | | | | | | | | | | | | | | |
Db 623 WTTAKAAATTTWKAATTTTGAAATW 598
```

Search completed: September 2, 2000, 00:24:41  
Job time: 4305 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2000, 23:40:36 : Search time 111.97 seconds  
(without alignments)  
1614.257 Million cell updates/sec

Title: US-09-464-528-6  
Perfect score: 1314  
Sequence: 1 tctagatcaaacctcacatcc.....ttttgaagtataaccatgg 1314

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/PCRTUS\_COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	5.9	7218	1	US-08-232-463-14
2	56	4.3	19124	4	US-08-487-826B-13
3	54.2	4.1	1186	4	US-08-731-722-5
4	53.6	4.1	19124	4	US-08-487-826B-13
5	52.6	4.0	665	4	US-08-883-795A-36
6	52	4.0	8920	3	US-08-446-855A-1
7	51.2	3.9	3933	4	US-08-731-722-3
8	51.2	3.9	3933	4	US-08-731-722-3
9	51	3.9	665	4	US-08-883-795A-36
10	50.8	3.9	602	1	US-08-229-393-1
11	50.2	3.8	473	2	US-08-764-100-16
12	50.2	3.8	4970	2	US-08-764-100-14
13	50.2	3.8	4970	2	US-08-764-100-20
14	50.2	3.8	8920	3	US-08-446-855A-1
15	49.8	3.8	4467	2	US-08-565-907A-1
16	49.8	3.8	4467	4	US-08-910-551B-1
17	49.8	3.8	4467	4	US-08-909-425A-1
18	49.4	3.8	1611	7	5213972-6
19	49.2	3.7	319	1	US-07-593-657-14
20	49	3.7	4098	4	US-08-605-106-4
21	48.2	3.7	6768	2	US-08-107-755A-1
22	48.2	3.7	8457	1	US-07-991-867B-1
23	48.2	3.7	8457	4	US-08-544-332-1
24	47.8	3.6	12124	1	US-08-181-271A-36
25	47.8	3.6	12124	1	US-08-449-315-36
26	47.8	3.6	12124	1	US-08-444-803-36

27 47.8 3.6 12124 1 US-08-449-043-36 Sequence 36, Appl  
28 47.8 3.6 12124 2 US-08-456-265A-36 Sequence 36, Appl  
29 47.8 3.6 12124 2 US-08-455-416-36 Sequence 36, Appl  
30 47.8 3.6 12124 2 US-08-455-244-36 Sequence 36, Appl  
31 47.8 3.6 12124 2 US-08-454-876-36 Sequence 36, Appl  
32 47.8 3.6 12124 3 US-08-457-364-36 Sequence 36, Appl  
33 47.8 3.6 12124 3 US-08-456-262-36 Sequence 36, Appl  
34 47.8 3.6 12124 3 US-08-456-240-36 Sequence 36, Appl  
35 47.8 3.6 12124 3 US-08-455-736-36 Sequence 36, Appl  
36 47.8 3.6 12124 4 US-08-971-217-36 Sequence 36, Appl  
37 47.6 3.6 10395 1 US-08-245-809-3 Sequence 3, Appl  
38 47.6 3.6 10396 1 US-08-245-809-5 Sequence 5, Appl  
39 47.6 3.6 10798 1 US-08-107-748-2 Sequence 2, Appl  
40 47.6 3.6 10798 6 PCT-US92-01385-2 Sequence 2, Appl  
41 47.6 3.6 10965 1 US-08-107-748-4 Sequence 4, Appl  
42 47.6 3.6 10965 6 PCT-US92-01385-4 Sequence 4, Appl  
43 47.4 3.6 642 2 US-08-764-100-13 Sequence 13, Appl  
44 47.4 3.6 643 2 US-08-764-100-7 Sequence 7, Appl  
45 47.4 3.6 2993 2 US-08-764-100-2 Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: ptzgt-F1s  
US-08-232-463-14



```
Query Match 5.9%; Score 78; DB 1; Length 7218;
Best Local Similarity 3.8%; Pred. No. 7.5e-07;
Matches 15; Conservative 241; Mismatches 136; Indels 0; Gaps 0;

Qy 554 cttttcgtcattaaactcaccgccaccggtttccctataaattggaactcaatgct 613
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1071 yvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv 1130

Qy 614 cccctctaaactcgtatcgcttcagagttgagacaaacacacactcgttcatactct 673
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1131 yvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv 1190

Qy 674 ctgctctctctctctctcctcctcctcaaggtactttctctcctcctcaccatccta 733
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1191 yvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv 1250

Qy 734 gattccggttgcaattcgatctgcaactctgtgttgccttgccttgccttgcctc 793
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1251 yvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv 1310

Qy 794 aactgggtccatctagatccatctgaaactcactcttcttctaatactcgcggaatc 853
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1311 yvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv 1370

Qy 854 gcgttggaactttcagatctagtcgaaatcatttcataatgccttcttcttgcctt 913
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1371 yvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv 1430

Qy 914 atgagaaataaaatcattttttttttttttttttttttttttttttttttttcaa 945
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1431 yvvvgtacaaattctctatctcttcttcttcttcttcttcttcttcttcttctt 1462

RESULT 2
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellem, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487.826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
```

```
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 4.3%; Score 56; DB 4; Length 19124;
Best Local Similarity 47.7%; Pred. No. 0.012;
Matches 164; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

Qy 57 attatttgggtggttaataatcattatttttaagatatttaataaagaaatataaagat 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1264 ATTATAATATGTAATATTATAATAATATATTCTTATAACATACAAACATAAAGAAAC 1323

Qy 117 tttttaaaaaatgtataaaattattatttatttgcatttgcatttgcatttgcatttgcata 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1324 TATACAATCTGCTATCTAATAGTATATATATATAATATCTTTTATTATTAATTGTTCTCT 1383

Qy 177 ataaatatatttttttttttttttttttttttttttttttttttttttttttttttttttt 236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1384 CTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1443

Qy 237 tggctgtttcacaaagcattcatttataacatttataacatttataacatttataacatttata 296
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1444 TTAGTATTATTATAATAATAATCTTTTAAAAAACTTCAMAAACATTTTGCATATAATA 1503

Qy 297 agaattcttctgtgagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 356
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1504 ATATTAATATTAGTAACCCACCTAGATAAATTAGAGAGAAACGTAGAACATACCAAAAAA 1563

Qy 357 gtcagacacagagacaaataaaagtgatgcaacaaacaaatca 400
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1564 ATTAGACAAAAAGAAATATTACAAAAAATAATAAAATTAATAATTA 1607

RESULT 3
US-08-731-722-5/C
; Sequence 5, Application US/08731722
; Patent No. 5961971
; GENERAL INFORMATION:
; APPLICANT: Martin, Frank N.
; TITLE OF INVENTION: Biocontrol of Fungal Soilborne Pathogens
; TITLE OF INVENTION: by Pythium oligandrum
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/731.722
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 1186 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: 23-5  
US-08-731-722-5

Query Match 4.13; Score 54.2; DB 4; Length 1186;  
Best Local Similarity 59.34; Pred. No. 0.018;  
Matches 112; Conservative 0; Mismatches 73; Indels 4; Gaps 1;

Qy 52 tactaattatttgggtaataatcattatttttaagataattaataagaattaa 111  
Db 595 TATTGACAAATTTCTTAACACTACTATTAAGATTTTAGGACTACTGTAATAATCCAG 536  
Qy 112 aagatttttaaaaaatgtataaaattatattattcatgattttttcatcatttgattt 171  
Db 535 ATCATTTTAAATAATAATAAAAGTTT---TTCAATATTATAATAAAATATATTC 480  
Qy 172 tgatacaataatatttttttttaattttcttaaaaaatgttgcaagacactcattagacat 231  
Db 479 TTATTAGAAGTATTTTCATCTTTAAATTTTAAAAAGTTATATATCTTTAAAAAGATAT 420  
Qy 232 agtcctgtt 240  
Db 419 AATTAAT 411

## RESULT 4

US-08-487-826B-13/C  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellem, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israel, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 4.13; Score 53.6; DB 4; Length 19124;  
Best Local Similarity 45.98; Pred. No. 0.033;  
Matches 221; Conservative 0; Mismatches 259; Indels 2; Gaps 1;

Qy 818 tgaactctactcttcttcttaatactcgcgaatacgcgttggacttccagatcagtcg 877  
Db 15940 TTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15881  
Qy 878 aaatcattcattcattcattcattcattcattcattcattcattcattcattcattcatt 937  
Db 15880 AATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15821  
Qy 938 tattcaaaaataaaccttggccttgcctgactgagatgggggttgggtgattacagaaat 997  
Db 15820 TATTTT---AATAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15763  
Qy 998 tttagcgaatttgcgaatttgcgaatttgcgaatttgcgaatttgcgaatttgcgaatttgc 1057  
Db 15762 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15703  
Qy 1058 catactccttagcttcaatttttattcagatagatagatagatagatagatagatagat 1117  
Db 15702 TTTATGATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15643  
Qy 1118 gagcagggggaatttaacccctccctccctccctccctccctccctccctccctccctcc 1177  
Db 15642 TTTTATTAATCATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15583  
Qy 1178 gaaacttttgccttcaaatctcattcattcattcattcattcattcattcattcattcatt 1237  
Db 15582 TCATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15523  
Qy 1238 ttgctctcctgtgtaaaatttactgttttaggtactaactaactaactaactaactaacta 1297  
Db 15522 TTTTCTCTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15463  
Qy 1298 tt 1299  
Db 15462 TT 15461

## RESULT 5

US-08-883-795A-36  
Sequence 36, Application US/08883795A  
Patent No. 5985607  
GENERAL INFORMATION:  
APPLICANT: Delcuve, Genevieve  
APPLICANT: Awang, Gregor  
TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
Vectors for Tissue Plasminogen Activator  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERESKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3J2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/883,795A  
FILING DATE: 27-JUN-1997

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gravelle, Micheline
REGISTRATION NUMBER: 40,361
REFERENCE/DOCKET NUMBER: 7841-062
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 665 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: Rh 32
US-08-883-795A-36

```

[illegible]

```

RESULT      6
US-08-446-855A-1/c
; Sequence 1, Application US/08446855A
; Patent No. 5849573
;
; GENERAL INFORMATION:
;
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence
; TITLE OF INVENTION: Nucleotide sequence
; TITLE OF INVENTION: phosphate synthe
;
; NUMBER OF SEQUENCES: 2
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Glebe RC
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
;
; ZIP: 22201-4714
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29.009

```

```

, REFERENCE/DOCKET NUMBER: 47-80
,
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: 703-816-4000
, TELEFAX: 703-816-4100
, INFORMATION FOR SEQ ID NO: 1:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 8520 base pairs
, TYPE: nucleic acid
, STRANDEDNESS: single
, TOPOLOGY: linear
, MOLECULE TYPE: genomic
US-08-446-855A-1

```

[illegible]

RESULT 7  
 US-08-731-722-3  
 ; Sequence 3, Application US/08731722  
 ; Patent No. 5961971  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Martin, Frank N.  
 ; TITLE OF INVENTION: Biocontrol of Fungal Soilborne Pathogens  
 ; TITLE OF INVENTION: by Pythium oligandrum  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Saliwanchik & Saliwanchik  
 ; STREET: 2421 N.W. 41st Street, Suite A-1  
 ; CITY: Gainesville  
 ; STATE: FL  
 ; COUNTRY: US  
 ; ZIP: 32606-6669  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/731,722  
 ; FILING DATE:  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Whitlock, Ted W.  
 ; REGISTRATION NUMBER: 36,965  
 ; REFERENCE/DOCKET NUMBER: UF-161  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 352-375-8100  
 ; TELEFAX: 352-372-5800  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:

LENGTH: 3933 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: 1986-41  
US-08-731-722-3

Query Match 3.9%; Score 51.2; DB 4; Length 3933;  
Best Local Similarity 54.3%; Pred. No. 0.076;  
Matches 125; Conservative 0; Mismatches 103; Indels 2; Gaps 1;

QY 53 actaatttttgggttaaaataatcaatcatttttttaagagataatcaatcaagaataaa 112  
DB 505 ATTATTATTCTTTTAAATAATATTATGATTATTATTAAAGTAGAATAAGG--ATTACA 562  
QY 113 agatttttttaaaaaatgtataaaattattattcattcattgatttttcacacattgatttt 172  
DB 563 ATTTTCTGTAAAAAATAAAACCTTCTCTATTGAGAAATTTTATTAGCTACTGTTTA 622  
QY 173 gataataatatttttttttttaatttttcaaaaatgttgcaagacacttattagacata 232  
DB 623 ATTAATAAAGTTATAATTAATTTCTCTATTGAGACATTTAAAAAATAATATTATAATG 682  
QY 233 gcttggtctgtttcaaaaagcattcattcatttataacattataaaataat 282  
DB 683 GTAGAATTGAGAGAATAAATACTTCTAATTACTTTATTATTATAATAT 732

## RESULT 8

US-08-731-722-3/c  
Sequence 3, Application US/08731722  
Patent No. 5961971

GENERAL INFORMATION:  
APPLICANT: Martin, Frank N.  
TITLE OF INVENTION: Biocontrol of Fungal Soilborne Pathogens  
TITLE OF INVENTION: by Pythium oligandrum  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL

COUNTRY: US  
ZIP: 32606-6669  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08731,722  
FILING DATE:

CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:

NAME: Whitlock, Ted W.  
REGISTRATION NUMBER: 36,965  
REFERENCE/DOCKET NUMBER: UF-161  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 352-375-8100  
TELEFAX: 352-372-5800

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 3933 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: 1986-41

US-08-731-722-3

Query Match 3.9%; Score 51.2; DB 4; Length 3933;  
Best Local Similarity 54.3%; Pred. No. 0.076;  
Matches 125; Conservative 0; Mismatches 103; Indels 2; Gaps 1;

QY 53 actaatttttgggttaaaataatcaatcatttttttaagagataatcaatcaagaataaa 112  
DB 3429 ATTATTATTCTTTTAAATAATATTATGATTATTATTAAAGTAGAATAAGG--ATTACA 3372  
QY 113 agatttttttaaaaaatgtataaaattattattcattcattgatttttcacacattgatttt 172  
DB 3371 ATTTTCTGTAAAAAATAAAACCTTCTCTATTGAGAAATTTTATTAGCTACTGTTTA 3312  
QY 173 gataataatatttttttttttaatttttcaaaaatgttgcaagacacttattagacata 232  
DB 3311 ATTAATAAAGTTATAATTAATTTCTCTATTGAGACATTTAAAAAATAATATTATAATG 3252  
QY 233 gcttggtctgtttcaaaaagcattcattcatttataacattataaaataat 282  
DB 3251 GTAGAATTGAGAGAATAAATACTTCTAATTACTTTATTATTATAATAT 3202

## RESULT 9

US-08-883-795A-36/c  
Sequence 36, Application US/08883795A  
Patent No. 5985607

GENERAL INFORMATION:  
APPLICANT: Delcuve, Genevieve  
APPLICANT: Awang, Gregor  
TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERESKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/883,795A  
FILING DATE: 27-JUN-1997

## CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Gravelle, Micheline  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 7841-062  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:  
LENGTH: 665 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: Rh 32

US-08-883-795A-36

Query Match 3.9%; Score 51; DB 4; Length 665;  
Best Local Similarity 53.0%; Pred. No. 0.066;  
Matches 131; Conservative 0; Mismatches 115; Indels 1; Gaps 1;





RESULT 14  
US-08-446-855A-1  
Sequence 1, Application US/08446855A  
Patent No. 5849573  
GENERAL INFORMATION:  
APPLICANT: Stewart, Thomas S  
APPLICANT: Flores, Maria V  
APPLICANT: O'Sullivan, William J  
TITLE OF INVENTION: Nucleotide sequence  
TITLE OF INVENTION: phosphate synthe  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRES:  
ADDRESSEE: Nixon & Vanderhye PC  
STREET: 1100 No. 5849573th Giebe-R  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM pc compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,855A  
FILING DATE: 06-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mitchard, Leonard C  
REGISTRATION NUMBER: 29 009  
REFERENCE/DOCKET NUMBER: 47-80  
TELEPHONE: 703-816-4000  
TELEFAX: 703-816-4100  
INFORMATION FOR SEQ ID NO: 1:

RESULT 45  
US-08-565-907A-1/c  
; Sequence 1, Application US/08565907A  
; Patent No. 5814499  
; GENERAL INFORMATION:  
; APPLICANT: Sylvain Moineau, Barbara  
; APPLICANT: J. Holler, Peter A. Vandenberg,  
; APPLICANT: Ebenezer R. Vedamuthu, Jeffrey K.  
; APPLICANT: Kondo  
; TITLE OF INVENTION: DNA Encoding Phage  
; TITLE OF INVENTION: Abortive Infection Protein  
; TITLE OF INVENTION: From Lactococcus  
; TITLE OF INVENTION: lactis, and Method of Use Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ian C. McLeod  
; STREET: 2190 Commons Parkway  
; CITY: Okemos  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 48864  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette 5.25 inch,  
; MEDIUM TYPE: 360 Kb storage  
; COMPUTER: Acer  
; OPERATING SYSTEM: MS-DOS (version 4)  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/565,907A  
; FILING DATE: December 1, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ian C. McLeod  
; REGISTRATION NUMBER: 20,931  
; REFERENCE/DOCKET NUMBER: Quest 4.1-152  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (517) 347-4100  
TELEFAX: (517) 347-4103  
TELEX: NO. 5814499e  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4467  
TYPE: Nucleotide  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE:  
DESCRIPTION: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N/A  
ORIGINAL SOURCE:  
ORGANISM: Lactococcus lactis  
STRAIN:  
INDIVIDUAL ISOLATE: W1  
DEVELOPMENTAL STAGE: N/A  
HAPLOTYPE: N/A  
TISSUE TYPE: N/A  
CELL TYPE: bacterium  
CELL LINE: N/A  
ORGANELLE: N/A  
IMMEDIATE SOURCE:  
LIBRARY: genomic  
CLONE: SMQ-20  
POSITION IN GENOME: N/A  
FEATURE:  
NAME/KEY: phage abortive infection  
LOCATION: N/A  
IDENTIFICATION METHOD: sequencing  
OTHER INFORMATION: DNA encoding phage  
OTHER INFORMATION: resistance  
PUBLICATION INFORMATION: N/A  
US-08-565-907A-1

Query Match 3.88; Score 49.8; DB 2; Length 4467;  
Best Local Similarity 49.48; Pred. No. 0.14; Mismatches 0; Gaps 0;  
Matches 129; Conservative 0; Indels 132; Gaps 0;  
Qy 51 atactaattatttgggttaataatcattatttttaagatatataaagaattta 110  
Db 2679 ATTTAAATCAATTAGTTAGTATTCTTCTGATTTTAAATCATCTATTTCAGAC 2620  
Qy 111 aaagattttttaaaaaatgataaaattatattattcattgatttttcacatttgatt 170  
Db 2619 AACCATATACAAATATTTCATATAATCTTGATTAATAATTTTTCGATAGTGGTC 2560  
Qy 171 ttgataataataatatt 230  
Db 2559 TATTTTTCCTTTTAAAGCCCTTGAATTTTACTAAATATTTTTCACATATTGAAAGC 2500  
Qy 231 tagcttctgtctgtttacaagcattcatcattataacattataaaatttttaataact 290  
Db 2499 TGATAAATCTGAATCAATTCATTAATTTTCAAGAAAGTCAAAACTTATTAGT 2440  
Qy 291 aacagtagaattctctgtga 311  
Db 2439 TAATCTCAATCTTTTAATGA 2419

Search completed: September 2, 2000, 01:57:31  
Job time: 8215 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2000, 23:33:51 ; Search time 161.21 Seconds  
(without alignments)  
2039.280 Million cell updates/sec

**Title:** US-09-464-528-6  
**perfect score:** 1314  
**Sequence:** 1 tctaatcaactcacatcc.....ttttgaagtataaacatga 1314

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

```
Maximum DB seq length: 9
Maximum DB seq length: 1000000
```

Post-processing: Minimum Match 08

Maximum Match 100%

**Listing first 45 summaries**

Database : N\_Geneseq\_36:★

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
C 1	65.8	5.0	6644	1	X33181	Base sequence of t
C 2	65.8	5.0	7372	1	X33182	Base sequence of t
C 3	65.8	5.0	7797	1	X33180	Compox virus bsr f
C 4	65.8	5.0	7996	1	X33184	Base sequence of t
C 5	60.2	4.6	9789	1	T41852	cDNA encoding Plas
C 6	56.4	4.3	5849	1	V33135	Plasmodium berghei
C 7	56	4.3	4590	1	N60472	Sequence encoding
C 8	56	4.3	19124	1	T72882	Plasmodium var-7 g
C 9	54.2	4.1	1186	1	V29580	Pythium oligandrum
C 10	53.6	4.1	19124	1	T72882	Plasmodium var-7 g
C 11	53.2	4.0	53585	1	X20251	Borrelia burgdorfe
C 12	52.6	4.0	605	1	T31530	Human 3' apolipop
C 13	52.4	4.0	26811	1	X20253	Borrelia burgdorfe
C 14	52.2	4.0	1826	1	V37413	Orpinomyces celluli
C 15	52.2	4.0	1826	1	V2947	Orpinomyces celluli
C 16	52.2	4.0	2503	1	V3480	pNPX30 xylanase cD
C 17	52	4.0	8920	1	Q62924	Carbamoyl-phosphat
C 18	51.8	3.9	605	1	T31530	Human 3' apolipop
C 19	51.8	3.9	3975	1	N81157	Malaria-specific g
C 20	51.8	3.9	92999	1	Q29999	SERP gene. Recombi
C 21	51.8	3.9	6124	1	Q03568	Sequence encoding
C 22	51.8	3.9	26811	1	X20253	Borrelia burgdorfe
C 23	51.4	3.9	9789	1	T41852	cDNA encoding Plas
C 24	51.2	3.9	783	1	X20361	Borrelia burgdorfe
C 25	51.2	3.9	1711	1	V33136	Plasmodium berghei
C 26	51.2	3.9	1864	1	N71405	Sequence of ANS-1
C 27	51.2	3.9	3933	1	V29578	Pythium oligandrum
C 28	51.2	3.9	3933	1	V29578	Pythium oligandrum
C 29	51	3.9	1470	1	O55185	MS-Le1610 Vector.
C 30	50.8	3.9	602	1	T58577	Cryptosporidium pa
C 31	50.8	3.9	110000	1	V21209.11	Continuation (12 o
C 32	50.2	3.8	2418	1	T27886	P. falciparum GBP13
C 33	50.2	3.8	4970	1	Q49959	Impatiens Necrotic

RESULT	1
--------	---

X33181/c  
 ID X33181 standard; DNA; 6644 BP.  
 AC X33181;  
 DT 25-JUN-1999 (first entry)  
 DE Base sequence of the plasmid pRX-ires  
 KW Cowpox virus; bsr: viral vector; expr  
 KW cma; bcl-2; bcl-x1; FLIP; survivin; r  
 KW autoimmune disease; graft rejection  
 KW inflammatory disease; ss.  
 OS Synthetic.  
 QS Cowpox virus.  
 OS Cowpox virus.  
 PN W0913073-A2.  
 PD 18-MAR-1999.  
 PF 07-SEP-1998; J04010.  
 PR 08-SEP-1997; JP-259235.  
 PG (RPRG-) RPR GENCELL ASIA PACIFIC INC.  
 PI Hamada H;  
 PI

Query Match 5.0%; Score 65.8; DB 1; Length 6644;  
Best Local Similarity 43.7%; Pred. No. 0.0045;  
Matches 289; Conservative 0; Mismatches 372; Indels 0

[illegible]









```

FT /number= 2
FT /note= "no stop codon given"
PN WO9640766-A2.
PD 19-DEC-1996.
PF 07-JUN-1996; U09508.
PR 07-JUN-1996; US-487826.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellens TE;
DR WPI: 97-052231/05.
DR P-PSDB; W22475.
PT New malaria vaccines - contains cysteine-rich DBL family protein
PT binding domains homologous domains of the Duffy and sialic acid
PT binding proteins
PS Claim 4; Page 56-61; 96pp; English.
CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to the
CC Duffy antigen binding protein (DABP) and sialic acid binding protein
CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC family of genes modulate cytoadherence and antigenic variation of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and SABP
CC mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. This sequence can be used in the compositions of the invention.
CC The compositions are for the treatment and prevention of malaria, and
CC comprise either a nucleotide sequence or encoded polypeptide of the
CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC genes having homology with conserved regions of DABP and SABP. The
CC compositions are used for the treatment and prevention of malaria. They
CC are also used in the preparation of vaccines for inducing a protective
CC immune response in a mammal to Plasmodium merozoites (especially
CC Plasmodium falciparum or Plasmodium vivax).
CC Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T;
SQ

```

```

KW soil saprophyte; pathogen; tomato; mycotoxic; plant protection;
KW vegetable crop; ss.
OS Pythium oligandrum.
FH Key Location/Qualifiers
FT misc_feature 502..684
FT /*tag= a
FT /note= "unique fragment claimed in claim 19"
PN WO9816110-A1.
PD 23-APR-1998.
PF 10-OCT-1997; U18343.
PR 17-OCT-1996; US-731722.
PA (UYFL ) UNIV FLORIDA.
PI Martin FN;
DR WPI: 98-250977/22.
PT Controlling phytopathogenic organisms with non-pathogenic Pythium
PT isolate - for control of damping off caused by Pythium
PS Claim 17; Page 30; 41pp; English.
CC This is a partial nucleotide sequence of the mitochondrial DNA from
CC a Pythium oligandrum isolate 23-5. Nucleotide sequences which are
CC inverted repeats, flanked by PstI restriction sites from mitochondrial
CC DNA from various P. oligandrum isolates are shown in V29576 to V29583.
CC These Pythium isolate sequences are non-pathogenic and can be used in a
CC method for controlling phytopathogenic organisms where the organisms are
CC contacted with such a Pythium isolate. The Pythium isolates are used to
CC control fungi, specifically pathogenic Pythium species (but possibly also
CC other soil-borne pathogens), particularly for protecting plants
CC (seedlings, transplants or vegetable crops such as tomato) against
CC damping off, especially after transplanting into open fields. The
CC isolate sequences are useful as sources of probes for identification of
CC particular isolates. The isolates are widely distributed in nature with
CC a similar ecology to pathogenic species. They produce large quantities of
CC oospore inoculum on liquid or solid substrates and are tolerant of
CC several commonly used fungicides. A single application at the greenhouse
CC stage will protect plants after transplanting.
CC Sequence 1186 BP; 469 A; 121 C; 125 G; 471 T;
SQ

```

```

Query Match 4.3%; Score 56; DB 1; Length 19124;
Best Local Similarity 47.7%; Pred. No. 0.17;
Matches 164; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 57 attatttgggttaatttaatttatttttaagatatttaagaaataaagat 116
DB 1264 ATTATTAATGTAATTAATTAATAAAATATATTGTATCAACATCAAGACTAAAGAAAC 1323
QY 117 ttttttaaaaaatgtataaaattattattcattgattttttcattcattgatttgata 176
DB 1324 TATCAACTGCTACTAATAGTATATATATATATATCTTTTATTATTAATGTCCT 1383
QY 177 atcaatatttttttttttttttttttttttttttttttttttttttttttttttttttttt 236
DB 1384 CTTTTTTTTTTTTTTTAAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1443
QY 237 tttttgttttttttttttttttttttttttttttttttttttttttttttttttttttttt 296
DB 1444 TTAGTATTTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1503
QY 297 aqaattcttcttgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 356
DB 1504 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1563
QY 357 gtcagacacagagacacacacacacacacacacacacacacacacacacacacacacacacac 400
DB 1564 ATTAGAACAACAAAGAAATATTACAAAAAATAAATAAATAAATAAATAAATAAATAAATA 1607

```

## RESULT 9

```

ID V29580/c
AC V29580;
DE Pythium oligandrum isolate 23-5 mitochondrial DNA partial sequence.
KW Pythium oligandrum; phytopathogenic; mitochondrial DNA; fungus;

```

```

Query Match 4.1%; Score 54.2; DB 1; Length 1186;
Best Local Similarity 59.3%; Pred. No. 0.33;
Matches 112; Conservative 0; Mismatches 73; Indels 4; Gaps 1;

QY 52 tactaatttttgggttaatttaatttatttttaagatatttaagaaataa 111
DB 595 TATTGAAGAATTTCTTATAAATCTACTATTAAAGATTATAGGATCTACTGTATAAATCCAG 536
QY 112 aagatttttttaaaaaatgtataaaattattattcattgattttttcattcattgatttgatt 171
DB 535 ATCATATTATAAATAAATAAATAAAGTTT----TTCAATATTATATAAATAATATTC 480
QY 172 tgataataataatatttttttttttttttttttttttttttttttttttttttttttttttt 231
DB 479 TTATTAGAAGTATTTCATTTTCAATTTTAAATTTTTTTTAAAGTTATATATATCTTTAAAGATAT 420
QY 232 agctctgtt 240
DB 419 AAATTAAT 411

```

## RESULT 10

```

ID T72882/c
AC T72882 standard; cDNA; 19124 BP.
DE T72882;
DE 12-SEP-1997 (first entry)
DE Plasmodium var-7 gene.
KW DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
KW Plasmodium; ss.
OS Plasmodium falciparum.
FH Key Location/Qualifiers
FT exon 7317..15139
FT /*tag= a

```

FT Intron /number= 1  
 FT 15140..16205  
 FT /tag= b  
 FT /number= 1  
 FT 16206..17552  
 FT /tag= c  
 FT /number= 2  
 FT /note= "no stop codon given"  
 PN W09640766-A2.  
 PD 19-DEC-1996.  
 PR 07-JUN-1995; U09508.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PI Chitnals C, Miller LH, Peterson DS, Sim KL, Su X;  
 PI Wellens TE;  
 DR WPI; 97-052231/05.  
 DR P-PSDB; W22475.  
 PT New malaria vaccines - contains cysteine-rich DBL family protein  
 PT binding domains homologous domains of the Duffy and sialic acid  
 PT binding proteins  
 PS Claim 4; Page 56-61; 96pp; English.  
 CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to  
 CC the Duffy binding like (DBL) family of genes which have homology to the  
 CC Duffy antigen binding protein (DABP) and sialic acid binding protein  
 CC (SABP) conserved regions (see T72889 and T72888 respectively). The var  
 CC family of genes modulate erythrocyte adherence and antigenic variation of  
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding  
 CC protein (DABP) are soluble proteins that appear in the culture  
 CC supernatant after infected erythrocytes release merozoites. DABP and SABP  
 CC mediate the binding of merozoites and schizonts to the erythrocyte  
 CC surface. These proteins are necessary for erythrocyte invasion by the  
 CC parasite. This sequence can be used in the compositions of the invention.  
 CC The compositions are for the treatment and prevention of malaria, and  
 CC comprise either a nucleotide sequence or encoded polypeptide of the  
 CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of  
 CC genes having homology with conserved regions of DABP and SABP. The  
 CC compositions are used for the treatment and prevention of malaria. They  
 CC are also used in the preparation of vaccines for inducing a protective  
 CC immune response in a mammal to Plasmodium merozoites (especially  
 CC Plasmodium falciparum or Plasmodium vivax).  
 CC Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T;  
 SQ

Query Match 4.1%; Score 53.6; DB 1; Length 19124;  
 Best Local Similarity 45.9%; Pred. No. 0.41;  
 Matches 221; Conservative 0; Mismatches 259; Indels 2; Gaps 1;

Qy 818 tgaacctactctcttcaataatcgcggaatacgcgttgagcattcagatcgtcg 877  
 Db 15940 TTAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTA 15881  
 Qy 878 aaatcatttcataatgccttcttcttcttcttcttcttcttcttcttcttctt 937  
 Db 15880 AATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15821  
 Qy 938 tattcacaataaaccttgcgcgttgcgcgttgcgcgttgcgcgttgcgcgttgcgc 997  
 Db 15820 TATTTT-AAATAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15763  
 Qy 998 tttagcgaatttgaattgactgttgcgttgcgttgcgttgcgttgcgttgcgttgc 1057  
 Db 15762 TTTTATTTTATTTATGATATATATTTTATTTTATTTTATTTTATTTTATTTT 15703  
 Qy 1058 catactcccttagcttcaatttcttcttcttcttcttcttcttcttcttcttct 1117  
 Db 15702 TTTATGATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15643  
 Qy 1118 gacgaggggaattaatcccttcccttcccttcccttcccttcccttcccttccct 1177  
 Db 15642 TTTTATAATCATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15583  
 Qy 1178 gaaacttttgccttaaatctcattataaacttttcttcttcttcttcttcttcttct 1237

Db 15582 TCATTTTATTTCTATCAAAATTTATATTTTATTTATTTTATTTTATTTTATTT 15523  
 Qy 1238 ttgctctctctgttaaaatttactgtttaggtactactactactgttctgttgcagttt 1297  
 Db 15522 TTTTCCTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15463  
 Qy 1298 tt 1299  
 Db 15462 TT 15461

RESULT 11  
 X20251/c  
 ID X20251 standard; DNA; 53585 BP.  
 AC X20251;  
 DT 04-MAY-1999 (first entry)  
 DE Borrelia burgdorferi polynucleotide sequence #4.  
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;  
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
 KW infection; diagnosis; characterisation; detection; ds.  
 OS Borrelia burgdorferi.  
 PN W09858943-AL.  
 PD 30-DEC-1998.  
 PF 18-JUN-1998; U12764.  
 PR 03-SEP-1997; US-057483.  
 PR 20-JUN-1997; US-050359.  
 PR 22-JUL-1997; US-053344.  
 PR 22-JUL-1997; US-053377.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMUNE INC.  
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,  
 PI White OR.  
 DR WPI; 99-081217/07.  
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of infections, particularly Lyme disease  
 PS Claim 1; Page 801-831; 1128pp; English.  
 CC X20248 to X20402 represent polynucleotide sequences isolated from  
 CC Borrelia burgdorferi (fb). Products derived from Bb can be used for  
 CC the detection, diagnosis, characterisation, prevention and therapy of  
 CC Bb infections, e.g. Lyme disease. They can also be used for the  
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs  
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
 CC Lyme disease.  
 CC Sequence 53585 BP; 19235 A; 8081 C; 7001 G; 19266 T;  
 SQ

Query Match 4.0%; Score 53.2; DB 1; Length 53585;  
 Best Local Similarity 52.2%; Pred. No. 0.48;  
 Matches 118; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy 21 aaacataacatggtatctctcttaccatcactactatttgggttaaatatcaatc 80  
 Db 33384 AAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 33325  
 Qy 81 attatttttaagatkatkaatttaagaataaagaattttttaaataaataaataa 140  
 Db 33324 CAAATTTTGTAAAAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTA 33265  
 Qy 141 tattattcattgatttttcatacatttgcatttgcataataataataataataattct 200  
 Db 33264 TTTTAAAGTTTATTTTGTGAAAAATATTATTGATGATGATTAATTAATTAATTA 33205  
 Qy 201 taaaaaatgttgcagacacatttagacatgcttcttcttcttcttcttcttctt 246  
 Db 33204 TTTTACCCTAAGGAGCTATTATGAAAAACAGATTTTCTCTATAT 33159

RESULT 12  
 T31530  
 ID T31530 standard; cDNA; 605 BP.

09-APR-1998. PD  
03-OCT-1997; U18008. PPF  
04-OCT-1996; US-027883. PPR  
(UYGE-) UNIV GEORGIA RES FOUND INC. PPA  
Chen H, Li X, Ljungdahl LG. PPI  
WPI; 98-240096/21. PPS  
P-PSDS; W56742. PDR  
New recombinant DNA encoding Orpinomyces cellulase protein - useful for, e.g. producing recombinant Orpinomyces cellulase in host cell Claim 5; Page 38-40; 69pp; English. PPT  
This cDNA clone includes a claimed coding region for cellulase celB (see W56742) of the anaerobic bovine rumen fungus Orpinomyces sp. PC-2. It was obtained by screening a PC-2 cDNA library for clones active on remazol brilliant blue-carboxymethylcellulose. The encoded cellulase has endoglucanase, but not cellobiohydrolase, activity. CelA and celC genes (see V29472-73), also obtained from Orpinomyces sp. PC-2, encode cellulases having both activities (see W56738-39). Recombinant DNA molecules encoding Orpinomyces cellulase proteins are claimed, as well as recombinant cells selected from Saccharomyces cerevisiae, Escherichia coli, Aspergillus, Trichoderma reesei, Pichia, Penicillium, Streptomyces



Query Match 4.0%; Score 52.2; DB 1; Length 1826;  
Best Local Similarity 49.8%;  
Pred. No. 0.69;  
Matches 132; Conservative 0; Mismatches 133; Indels 0

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2000, 23:32:16 ; Search time 5217.98 seconds  
(without alignments)  
449.425 Million cell updates/sec

Title: us-09-464-528-6  
Perfect score: 1314  
Sequence: 1 tctagataaacacacatcc.....ttttgaagtataaccatgg 1314

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

GenEmbl:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: em\_fun:\*  
17: em\_hum1:\*  
18: em\_hum2:\*  
19: em\_in:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_sy:\*  
29: em\_un:\*  
30: em\_v1:\*  
31: gb\_htg1:\*  
32: gb\_htg2:\*  
33: gb\_in1:\*  
34: gb\_in2:\*  
35: em\_ba1:\*  
36: em\_ba2:\*  
37: em\_hum3:\*  
38: em\_hum4:\*  
39: gb\_pr4:\*  
40: gb\_htg3:\*  
41: gb\_htg4:\*  
42: gb\_htg5:\*  
43: gb\_htg6:\*

44: gb\_htg7:\*  
45: em\_htg1:\*  
46: em\_htg2:\*  
47: em\_htg3:\*  
48: em\_hum5:\*  
49: gb\_pl3:\*  
50: gb\_pr5:\*  
51: gb\_htg8:\*  
52: gb\_htg9:\*  
53: gb\_htg10:\*  
54: gb\_htg11:\*  
55: gb\_htg12:\*  
56: gb\_htg13:\*  
57: gb\_htg14:\*  
58: gb\_in3:\*  
59: gb\_htg15:\*  
60: gb\_htg16:\*  
61: gb\_htg17:\*  
62: em\_htg4:\*  
63: em\_htg5:\*  
64: em\_htg6:\*  
65: em\_htg7:\*  
66: em\_hum6:\*  
67: gb\_htg18:\*  
68: gb\_htg19:\*  
69: gb\_htg20:\*  
70: gb\_htg21:\*  
71: gb\_htg22:\*  
72: gb\_htg23:\*  
73: gb\_htg24:\*  
74: gb\_htg25:\*  
75: gb\_htg26:\*  
76: gb\_htg27:\*  
77: gb\_htg28:\*  
78: gb\_htg29:\*  
79: gb\_htg30:\*  
80: gb\_htg31:\*  
81: gb\_v11:\*  
82: gb\_v12:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	5.9	7218	5	I66494 Sequence 14
2	74	5.6	144759	10	AL021939 Homo sapi
3	70.2	5.3	153267	57	AC011212 Homo sapi
4	70	5.3	104992	41	AC005504 Plasmodi
5	70	5.3	130281	60	AC004157 Plasmodi
6	69.8	5.3	349919	54	AC008576 Homo sapi
7	69.6	5.3	164119	75	AC026640 Homo sapi
8	69.2	5.3	179310	71	AC013820 Homo sapi
9	69	5.3	48532	51	AC023371 Homo sapi
10	69	5.3	204951	60	AC005505 Plasmodi
11	68.4	5.2	975	33	U87514 Dictyosteli
12	68.2	5.2	80920	60	AC006278 Plasmodi
13	68.2	5.2	121024	8	X04465 Marchantia
14	67.8	5.2	910	13	AL142826 Anopheles
15	67.4	5.1	256172	41	AC005139 Plasmodi
16	67.2	5.1	43346	55	AC011556 Homo sapi
17	66.8	5.1	99263	32	AL139177 Homo sapi
18	66.8	5.1	152209	11	AL034419 Human DNA
19	66.6	5.1	1867	7	AJ223323 Saccharom
20	66.4	5.1	149752	11	AC004616 Homo sapi
21	66.4	5.1	170427	41	AC006095 Homo sapi
22	66.4	5.1	183638	77	AC009653 Homo sapi
23	66.4	5.1	192581	31	AL049180 Plasmodi
24	66.4	5.1	216406	52	AC018919 Homo sapi

```
25 66.2 5.0 167271 53 AC006269
26 66.2 5.0 226345 39 AC005406
27 66 5.0 158131 68 AC009277
28 65.8 5.0 2982 33 DDISGSPA
29 65.8 5.0 67970 33 PMALIP3
30 65.8 5.0 68746 43 AC021008
31 65.8 5.0 162575 41 AC004086
32 65.4 5.0 690 7 MISC10
33 65.4 5.0 716 8 YSCMTTGT1
34 65.2 5.0 68879 72 AC031982
35 65.2 5.0 161891 55 AC008206
36 65.2 5.0 287731 59 AC008620
37 65 4.9 2305 33 DPMTRNA
38 65 4.9 69412 43 AC021521
39 65 4.9 129404 51 AC013349
40 64.8 4.9 163678 52 AC010178
41 64.8 4.9 173693 57 AC021553
42 64.6 4.9 52359 41 AC010772
43 64.4 4.9 209734 32 CNS01RG3
44 64.4 4.9 217242 73 AC016222
45 64.2 4.9 153267 57 AC011212

ALIGNMENTS

RESULT 1
LOCUS I66494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
source
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
Query Match 5.9%; Score 78; DB 5; Length 7218;
Best Local Similarity 3.8%; Pred. No. 0.072;
Matches 15; Conservative 241; Mismatches 136; Indels 0; Gaps 0;

QY 554 ctttcgctcaataactacccctgcacccggtttccctataaattggaactcaatgct 613
Db 1071 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1130
QY 614 ccccttaactcgtatcgttcagagttgagaccagacacactcgttcatactct 673
Db 1131 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1190
QY 674 ctgctctctctctctctacctcctcaagggtactttctctccctcacaaacctta 733
Db 1191 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1250
QY 734 gattccggtggtcaattcgggactctgcactctggttgccttgccttctctcctc 793
Db 1251 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1310
QY 794 aactcgttcactagatccatcgtgaactcactctcttcttcttcttctcgtggaatc 853
Db 1311 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1370
QY 854 gccttagcttcagatcgttcagatccattcattcattcattcattcattcattcatt 913
Db 913 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
```

```
Db 1371 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1430
QY 914 atgagaaataaatacatcttttttttttttttttttttttttttttttttttttt 945
Db 1431 YYYGGTACCAAAATCTCTATCTCTTAACTA 1462

RESULT 2
HS352A20 144759 bp DNA PRI 23-NOV-1999
LOCUS HS352A20
DEFINITION Homo sapiens DNA sequence from PAC 352A20 on chromosome 6q24.1-25.1. Contains a pseudogene similar to yeast, bacterial, worm and slime mold hypothetical genes, and a gene coding for an aldehyde dehydrogenase family protein. Contains ESTs, STSs and GSSs, complete sequence.
ACCESSION AL021939
VERSION AL021939.1 GI:3135969
KEYWORDS HTG: aldehyde dehydrogenase.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 144759)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL Phillips,S.
TITLE Direct Submission
COMMENT Submitted (07-MAY-1998) sanger.ac.uk/HGP/Chr6/ Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk On May 18, 1998 this sequence version replaced gi:2909620. IMPORTANT: This sequence is the entire insert of clone 352A20. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variations annotated may not be found in the sequence submission corresponding to the overlapping clone as we submit sequences with only a small overlap as described above. This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre chromosome 6 mapping group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6/ This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The true left end of clone 352A20 is at 1 in this sequence. The true right end of clone 352A20 is at 144759. 352A20 is from the library RPCI3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/.
FEATURES
source
Location/Qualifiers
1..144759
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q24.1-25.1"
/clone="RP3-352A20"
/clone_lib="RPCI-3"
repeat_region 3..211
/note="AluSc repeat: matches 77. .285 of consensus; incomplete repeat"
repeat_region 334..392
/note="MER45 repeat: matches 4. .62 of consensus"
repeat_region 662..959
/note="AluSc repeat: matches 1. .299 of consensus"
repeat_region 1362..1544
/note="LIPB3 repeat: matches 640. .811 of consensus"
repeat_region 1546..1836
/note="AluSc repeat: matches 299. .1 of consensus"
repeat_region 1839..1937
/note="L1M4A repeat: matches 941. .1046 of consensus"
repeat_region 2448..2745
```









```
* 56312 59177: contig of 2866 bp in length
* gap of unknown length
* 59178 61728: contig of 2551 bp in length
* gap of unknown length
* 61729 65138: contig of 3410 bp in length
* gap of unknown length
* 65139 68509: contig of 3371 bp in length
* gap of unknown length
* 68510 72140: contig of 3631 bp in length
* gap of unknown length
* 72141 74177: contig of 2037 bp in length
* gap of unknown length
* 74178 76657: contig of 2480 bp in length
* gap of unknown length
* 76658 81120: contig of 4463 bp in length
* gap of unknown length
* 81121 83944: contig of 2824 bp in length
* gap of unknown length
* 83945 88969: contig of 5025 bp in length
* gap of unknown length
* 88970 94016: contig of 5047 bp in length
* gap of unknown length
* 94017 99820: contig of 5804 bp in length
* gap of unknown length
* 99821 103739: contig of 3919 bp in length
* gap of unknown length
* 103740 110294: contig of 6555 bp in length
* gap of unknown length
* 110295 116891: contig of 6597 bp in length
* gap of unknown length
* 116892 122399: contig of 5508 bp in length
* gap of unknown length
* 122400 129404: contig of 7005 bp in length
* gap of unknown length
* 129405 136213: contig of 6809 bp in length
* gap of unknown length
* 136214 142142: contig of 5929 bp in length
* gap of unknown length
* 142143 150401: contig of 8259 bp in length
* gap of unknown length
* 150402 158946: contig of 8545 bp in length
* gap of unknown length
* 158947 167955: contig of 9009 bp in length
* gap of unknown length
* 167956 177918: contig of 9963 bp in length
* gap of unknown length
* 177919 183990: contig of 6072 bp in length
* gap of unknown length
* 183991 191939: contig of 7949 bp in length
* gap of unknown length
* 191940 201055: contig of 9116 bp in length
* gap of unknown length
* 201056 210412: contig of 9357 bp in length
* gap of unknown length
* 210413 219973: contig of 9561 bp in length
* gap of unknown length
* 219974 228547: contig of 8574 bp in length
* gap of unknown length
* 228548 252971: contig of 24424 bp in length
* gap of unknown length
* 252972 284497: contig of 31526 bp in length
* gap of unknown length
* 284498 316714: contig of 32217 bp in length
* gap of unknown length
* 316715 349919: contig of 33205 bp in length.
* Location/Qualifiers
* 1.349919
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="19"
* /clone="CTC-557J18"
* 96739 a 81094 c 78642 g 92256 t 1188 others
* BASE COUNT
* ORIGIN
```

```
Query Match 5.3%; Score 69.8; DB 54; Length 349919;
Best Local Similarity 57.2%; Pred. No. 0.26;
Matches 147; Conservative 0; Mismatches 107; Indels 3; Gaps 1;

QY 53 actaatattttgggttaaatatcatcatttttaagatat---taattaagaatt 109
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309365 ACAACAAATTCGTTAAATGAAATGTTTATAAATATATATATTAATATAT 309306

QY 110 aaagaatttttaaaaaatgataaattatattattcattcattttcattcattgat 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309305 TATATATATATTAATATATATATATATATATATATATATATATATATATAT 309246

QY 170 ttgtataataataatttttttaatttttaaaaaatgttgcagacacattattagac 229
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309245 TATAATATATATATATATATATATATATATATATATATATATATATATATAT 309186

QY 230 atagcttctgttttacaaaagcattcatcatttaacattacataaaattttaataac 289
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309185 TTATAATATATATATATATATATATATATATATATATATATATATATATATAT 309126

QY 290 taacagtagaattctct 306
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309125 ATTTATATATATATCT 309109

RESULT 7
AC026640 164119 bp DNA HTG 13-APR-2000
LOCUS Homo sapiens chromosome 11 clone RP11-69K18 map 11, WORKING DRAFT
DEFINITION SEQUENCE, 16 unordered pieces.
ACCESSION AC026640
VERSION AC026640.2 GI:7547222
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 164119)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abram,H., Allen,N.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehotzky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 13, 2000 this sequence version replaced gi:7284664.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
```



Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L8749

Center clone name: 69\_K\_18

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.96073

Consensus quality: 150998 bases at least Q40

Consensus quality: 158219 bases at least Q30

Consensus quality: 160955 bases at least Q20

Insert size: 176000; agarose-fp

Quality coverage: 162619; sum-of-contigs

Quality coverage: 3.8 in Q20 bases; agarose-fp

Quality coverage: 4.1 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1608: contig of 1608 bp in length
* 1609 1708: gap of 100 bp
* 1709 3184: contig of 1476 bp in length
* 3185 3284: gap of 100 bp
* 3285 5541: contig of 2257 bp in length
* 5542 5641: gap of 100 bp
* 5642 7773: contig of 2132 bp in length
* 7774 7873: gap of 100 bp
* 7874 10690: contig of 2817 bp in length
* 10691 10790: gap of 100 bp
* 10791 13433: contig of 2643 bp in length
* 13434 13533: gap of 100 bp
* 13534 16455: contig of 2922 bp in length
* 16456 16555: gap of 100 bp
* 16556 21573: contig of 5018 bp in length
* 21574 21673: gap of 100 bp
* 21674 25705: contig of 4032 bp in length
* 25706 25805: gap of 100 bp
* 25806 31983: contig of 6178 bp in length
* 31984 32083: gap of 100 bp
* 32084 42802: contig of 10719 bp in length
* 42803 42902: gap of 100 bp
* 42903 54889: contig of 11987 bp in length
* 54890 54989: gap of 100 bp
* 54990 67682: contig of 12693 bp in length
* 67683 67782: gap of 100 bp
* 67783 80005: contig of 12223 bp in length
* 80006 80105: gap of 100 bp
* 80106 106699: contig of 26594 bp in length
* 106700 106799: gap of 100 bp
* 106800 164119: contig of 57320 bp in length.
Location/Qualifiers
1. 164119
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-69K18"
/clone_lib="RPC1-11 Human Male BAC"
1. 1608
/organism="Homo sapiens"
/clone="assembly_fragment"
1709. 3184
/organism="Homo sapiens"
/clone="assembly_fragment"
3285. 5541
/organism="Homo sapiens"
/clone="assembly_fragment"
5542. 7773
/organism="Homo sapiens"
/clone="assembly_fragment"

```

# FEATURES

source

```

misc_feature 7874..10690
             /note="assembly_fragment"
misc_feature 10791..13433
             /note="assembly_fragment"
misc_feature 13534..16455
             /note="assembly_fragment"
misc_feature 16556..21573
             /note="assembly_fragment"
misc_feature 21674..25705
             /note="assembly_fragment"
misc_feature 25806..31983
             /note="assembly_fragment"
             clone_end:SP6
             vector side:left
misc_feature 32084..42802
             /note="assembly_fragment"
misc_feature 42903..54889
             /note="assembly_fragment"
misc_feature 54990..67682
             /note="assembly_fragment"
misc_feature 67783..80005
             /note="assembly_fragment"
misc_feature 80106..106699
             /note="assembly_fragment"
misc_feature 106800..164119
             /note="assembly_fragment"

```

BASE COUNT 51462 a 29898 c 30038 g 51215 t 1506 others  
ORIGIN

Query Match 5.3%; Score 69.6; DB 75; Length 164119;  
Best Local Similarity 55.3%; Pred. No. 0.35;  
Matches 135; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```

QY 68 ttaataataatcattttttaagataataaataaagaataaaagatttttaaaaaa 127
Db 118812 TAATATATATATATATATATATATATATATATATATATATATATATATAT 118871
QY 128 atgtataaaattatattcattcattttttcattcattttttttgataataataat 187
Db 118872 ATATATATATATATATATATATATATATATATATATATATATATATATAT 118931
QY 188 ttttttaatttttaaaaaattgttgcaagacactatttagacatagcttctgttta 247
Db 118932 TTAATTTTATATATATATATATATATATATATATATATATATATATATAT 118991
QY 248 caaagcattcattcatttaataacattaaaaaataatttaataactaacagtagaattctt 307
Db 118992 ATATATATATATATATATATATATATATATATATATATATATATATATAT 119051
QY 308 gtga 311
Db 119052 GAGA 119055

```

# RESULT

8

AC013820/c

LOCUS

DEFINITION

AC013820

VERSION

AC013820.3

KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 179510)

AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE

Homo sapiens, clone RP11-21P24.

JOURNAL

Unpublished.

REFERENCE

2 (bases 1 to 179510)

AC013820 179510 bp DNA HTG 01-APR-2000  
Homo sapiens clone RP11-21P24, WORKING DRAFT SEQUENCE, 8 unordered  
pieces.  
AC013820  
AC013820.3 GI:7382157  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 179510)  
AUTHORS  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE  
Homo sapiens, clone RP11-21P24.  
JOURNAL  
Unpublished.  
REFERENCE  
2 (bases 1 to 179510)



McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J.,  
 Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T. M.,  
 Peterson, K., Piere, N., Pisani, C., Pollara, V., Raymond, C.,  
 Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,  
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,  
 Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,  
 Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and  
 Zody, M.

## Direct Submission

Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIGR

Web site: <http://www-seq.wi.mit.edu>Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L3985

Center clone name: 21\_D\_18

-----

\* NOTE: This record contains 55 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1  
 869: contig of 869 bp in length  
 gap of unknown length  
 870 1715: contig of 846 bp in length  
 gap of unknown length  
 1716 2597: contig of 882 bp in length  
 gap of unknown length  
 2598 3468: contig of 871 bp in length  
 gap of unknown length  
 3469 4369: contig of 901 bp in length  
 gap of unknown length  
 4370 5239: contig of 870 bp in length  
 gap of unknown length  
 5240 6103: contig of 864 bp in length  
 gap of unknown length  
 6104 6975: contig of 872 bp in length  
 gap of unknown length  
 6976 7871: contig of 896 bp in length  
 gap of unknown length  
 7872 8739: contig of 868 bp in length  
 gap of unknown length  
 8740 9613: contig of 874 bp in length  
 gap of unknown length  
 9614 10472: contig of 859 bp in length  
 gap of unknown length  
 10473 11328: contig of 856 bp in length  
 gap of unknown length  
 11329 12188: contig of 860 bp in length  
 gap of unknown length  
 12189 13069: contig of 881 bp in length  
 gap of unknown length  
 13070 13952: contig of 883 bp in length  
 gap of unknown length  
 13953 14829: contig of 877 bp in length  
 gap of unknown length  
 14830 15701: contig of 872 bp in length  
 gap of unknown length  
 15702 16587: contig of 886 bp in length

16588 17461: contig of 874 bp in length  
 gap of unknown length  
 17462 18326: contig of 865 bp in length  
 gap of unknown length  
 18327 19191: contig of 865 bp in length  
 gap of unknown length  
 19192 20081: contig of 890 bp in length  
 gap of unknown length  
 20082 20971: contig of 890 bp in length  
 gap of unknown length  
 20972 21849: contig of 878 bp in length  
 gap of unknown length  
 21850 22713: contig of 864 bp in length  
 gap of unknown length  
 22714 23562: contig of 849 bp in length  
 gap of unknown length  
 23563 24422: contig of 860 bp in length  
 gap of unknown length  
 24423 25286: contig of 864 bp in length  
 gap of unknown length  
 25287 26173: contig of 887 bp in length  
 gap of unknown length  
 26174 27055: contig of 882 bp in length  
 gap of unknown length  
 27056 28036: contig of 981 bp in length  
 gap of unknown length  
 28037 28916: contig of 880 bp in length  
 gap of unknown length  
 28917 29800: contig of 884 bp in length  
 gap of unknown length  
 29801 30670: contig of 870 bp in length  
 gap of unknown length  
 30671 31669: contig of 999 bp in length  
 gap of unknown length  
 31670 32564: contig of 895 bp in length  
 gap of unknown length  
 32565 33435: contig of 871 bp in length  
 gap of unknown length  
 33436 34363: contig of 928 bp in length  
 gap of unknown length  
 34364 35264: contig of 901 bp in length  
 gap of unknown length  
 35265 36131: contig of 867 bp in length  
 gap of unknown length  
 36132 37024: contig of 893 bp in length  
 gap of unknown length  
 37025 37922: contig of 898 bp in length  
 gap of unknown length  
 37923 38829: contig of 907 bp in length  
 gap of unknown length  
 38830 39709: contig of 880 bp in length  
 gap of unknown length  
 39710 40590: contig of 881 bp in length  
 gap of unknown length  
 40591 41490: contig of 900 bp in length  
 gap of unknown length  
 41491 42372: contig of 882 bp in length  
 gap of unknown length  
 42373 43253: contig of 881 bp in length  
 gap of unknown length  
 43254 44133: contig of 880 bp in length  
 gap of unknown length  
 44134 45009: contig of 876 bp in length  
 gap of unknown length  
 45010 45881: contig of 872 bp in length  
 gap of unknown length  
 45882 46757: contig of 876 bp in length  
 gap of unknown length  
 46758 47645: contig of 888 bp in length  
 gap of unknown length  
 47646 48532: contig of 887 bp in length.  
 Location/Qualifiers

FEATURES

FEATURES	source	Location/Qualifiers
		1. .204951
		/organism="Plasmodium falciparum"
		/db_xref="taxon:5833"
		/chromosome="12"
		/clone="3D7"
BASE COUNT	83509 a	19825 c 24530 g 74582 t 2405 others
ORIGIN		
Query Match	5.3%	Score 69; DB '60; Length 204951;
Best Local Similarity	49.2%	Pred. No. 0.39; 215; Indels 1; Gaps 1;
Matches 209;	Conservative	0; Mismatches
QY 884	tttcataatgctcttcttcttttagcttatgagaaataaacatctttttttatttc 943	
Db 149838	TTTATTTAAATTAATTTTTTTTATTATTATTAATAAATTTTTTAATTTTTTTTATTTA 149779	
QY 944	aaataaaccttggcccttgctgctgactgagatggggcttggtgattacagaaatttagc 1003	
Db 149778	AATAAATTTTTTTTATTATTATTATTAATAAATTTTTTTTAGTTTTTTTATTTTTTTTA-A 149720	
QY 1004	gaatttgttaattgactgttctgctgtagtttctgttcttctgttctgttcttcataca 1063	
Db 149719	ATATTTTTTATTTATGATATATATTTTATTTTAAATATATTTTTTTTGGCTTTTTTTTAA 149660	
QY 1064	ttcctaggcttcaattttatttcgagtatagggtcacataggaaattcaaaccttgagcag 1123	
Db 149659	TTTTATGATATATATTTTTTTTTTAAATATATTTTTTCTTTTTTTTATTTTATGATAT 149600	
QY 1124	gggaattaaaccttcctccaaaccaggtttgttgatatatatgttttaaaaaaagaaact 1183	
Db 149599	ATATTTTTTTTTTAAATGTTTTTTTTCTCTCTTTTTTATTATTATTTCTAATAAATTT 149540	
QY 1184	ttgctttaaattctataaacttttttatacgcaaaaatttttgcattgctgtcttgcct 1243	
Db 149539	TTTTTATTTTTTTTAAATCAATTTTTTTTATATATAAATAATTTTTTTTAAATTCCTTTTG 149480	
QY 1244	ctcctgtgtaaaatttactgttttaggtactaacctcaggcttgttgtagcagttttggaag 1303	
Db 149479	ATAATATTTTTTTTATTTTTTATTATCAAAATTTATATTTTTTAAATTTTTTATTACT 149420	
QY 1304	tataa 1308	
Db 149419	TTTAA 149415	
RESULT 11		
DDU87514/c		
LOCUS	975 bp	DNA INV 05-APR-1997
DEFINITION	Dictyostelium discoideum CAR3 gene, promoter region.	
ACCESSION	U87514	
VERSION	U87514.1	GI:1927211
KEYWORDS	Dictyostelium discoideum.	
ORGANISM	Dictyostelium discoideum	



REFERENCE AUTHORS TITLE	1 (bases 1 to 121024) Yamano, Y., Ohshima, K. and Komano, T. Nucleotide sequences of chloroplast 5S ribosomal RNA from cell suspension cultures of the liverworts <i>Marchantia polymorpha</i> and <i>Jungmannia subulata</i> Nucleic Acids Res. 12 (11), 4621-4624 (1984) 84247325	TITLE JOURNAL MEDLINE REFERENCE AUTHORS	Structure and organization of <i>Marchantia polymorpha</i> chloroplast genome. I. Cloning and gene identification J. Mol. Biol. 203 (2), 281-298 (1988) 89068685
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	2 (bases 1 to 121024) Umesono, K., Inokuchi, H., Ohshima, K. and Ozeki, H. Nucleotide sequence of <i>Marchantia polymorpha</i> chloroplast DNA: a region possibly encoding three tRNAs and three proteins including a homologue of E. coli ribosomal protein S14 Nucleic Acids Res. 12 (24), 9551-9565 (1984) 85087956	TITLE JOURNAL MEDLINE REFERENCE AUTHORS	Structure and organization of <i>Marchantia polymorpha</i> chloroplast genome. III. Gene organization of the large single copy region from rbcL to trnI(CAU) J. Mol. Biol. 203 (2), 333-351 (1988) 89068687
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	3 (bases 1 to 121024) Fukuzawa, H., Uchida, Y., Yamano, Y., Ohshima, K. and Komano, T. Molecular cloning of promoters functional in <i>Escherichia coli</i> from chloroplast DNA of a liverwort, <i>Marchantia polymorpha</i> Agric. Biol. Chem. 49, 2725-2731 (1985)	TITLE JOURNAL MEDLINE REFERENCE AUTHORS	Shimada, H. and Sugiyama, M. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes Nucleic Acids Res. 19 (5), 983-995 (1991) 91212240
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	4 (bases 1 to 121024) Yamano, Y., Kohchi, T., Fukuzawa, H., Ohshima, K. and Komano, T. Nucleotide sequences of chloroplast 4.5 S ribosomal RNA from a leafy liverwort, <i>Jungmannia subulata</i> , and a thalloid liverwort, <i>Marchantia polymorpha</i> FEBS Lett. 185, 203-207 (1985)	TITLE JOURNAL MEDLINE REFERENCE AUTHORS	Direct Submission Submitted (17-OCT-1986) K. Ohshima, Research Centre for Cell and Tissue culture, Faculty of Agriculture, Kyoto University, Kyoto 606, Japan GENERAL DESCRIPTION OF SEQUENCE gene name product name 16S rRNA 23S rRNA Ribosomal RNA genes 4.5S rRNA 5S rRNA
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	5 (bases 1 to 121024) Ohshima, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. Chloroplast gene organization deduced from complete sequence of liverwort <i>Marchantia polymorpha</i> chloroplast DNA Nature 322, 572-574 (1986)	COMMENT	RNA polymerase genes rpoA: homologous to E. coli beta rpoB: homologous to E. coli beta rpoC1: homologous to E. coli beta rpoC2: homologous to E. coli Ribosomal protein genes and related genes 50S subunit rpl2: homologous to E. coli L2 L14 rpl14: homologous to E. coli L16 rpl16: homologous to E. coli L20 rpl20: homologous to E. coli L21 rpl21: homologous to E. coli L22 rpl22: homologous to E. coli L23 rpl23: homologous to E. coli L33 rpl33: homologous to E. coli 30S subunit rps2: homologous to E. coli S2 rps3: homologous to E. coli S3 rps4: homologous to E. coli S4 rps7: homologous to E. coli S7 rps8: homologous to E. coli S8 S11 rps11: homologous to E. coli S12 rps12: homologous to E. coli S14 rps14: homologous to E. coli S15 rps15: homologous to E. coli S18 rps18: homologous to E. coli S19 rps19: homologous to E. coli other genes homologous to E. coli Genes for photosynthesis rbcL: large subunit of Rubisco psaA: photosystem I P700 chlorophyll (an apoprotein) psbA: same as above psbB: photosystem II 32kd protein psbB: photosystem II P680 chlorophyll (an apoprotein) psbC: same as above psbD: photosystem II D2 protein psbE: cytochrome b559 psbF: same as above
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	6 (bases 1 to 121024) Fukuzawa, H., Kohchi, T., Shirai, H., Ohshima, K., Umesono, K., Inokuchi, H. and Ozeki, H. Coding sequences for chloroplast ribosomal protein S12 from the liverwort, <i>Marchantia polymorpha</i> , are separated far apart on the different DNA strands FEBS Lett. 198, 11-15 (1986)		
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	7 (bases 1 to 121024) Fukuzawa, H., Yoshida, T., Kohchi, T., Okumura, T., Sawano, Y. and Ohshima, K. Splicing of group II introns in mRNAs coding for cytochrome b6 and subunit IV in liverwort <i>Marchantia polymorpha</i> chloroplast genome: Exon specifying a region coding for two genes with the spacer region FEBS Lett. 220, 61-66 (1987)		
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	8 (bases 1 to 121024) Kohchi, T., Ogura, Y., Umesono, K., Yamada, Y., Komano, T., Ozeki, H. and Ohshima, K. Ordered processing and splicing in a polycistronic transcript in liverwort chloroplasts Curr. Genet. 14 (2), 147-154 (1988) 89028845		
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	9 (bases 1 to 121024) Kohchi, T., Shirai, H., Fukuzawa, H., Sano, T., Komano, T., Umesono, K., Inokuchi, H., Ozeki, H. and Ohshima, K. Structure and organization of <i>Marchantia polymorpha</i> chloroplast genome. IV. Inverted repeat and small single copy regions J. Mol. Biol. 203 (2), 353-372 (1988) 89068688		
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	10 (bases 1 to 121024) Umesono, K., Inokuchi, H., Shiki, Y., Takeuchi, M., Chang, Z., Fukuzawa, H., Kohchi, T., Shirai, H., Ohshima, K. and Ozeki, H. Structure and organization of <i>Marchantia polymorpha</i> chloroplast genome. II. Gene organization of the large single copy region from rps'12 to atp8 J. Mol. Biol. 203 (2), 299-331 (1988) 89068686		
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	11 (bases 1 to 121024) Ohshima, K., Fukuzawa, H., Kohchi, T., Sano, T., Sano, S., Shirai, H., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H.		







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 00:24:41 ; Search time 2111.72 Seconds  
(without alignments)  
4521.091 Million cell updates/sec

Title: US-09-464-528-14  
Perfect score: 2165  
Sequence: 1.atcgatagacatgttatt.....gttttgaagtataaagatg 2165

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*

44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: em\_est35:\*  
83: em\_est36:\*  
84: em\_est37:\*  
85: gb\_est48:\*  
86: gb\_est49:\*  
87: gb\_est50:\*  
88: gb\_est51:\*  
89: gb\_est52:\*  
90: gb\_est53:\*  
91: gb\_est54:\*  
92: gb\_est55:\*  
93: gb\_gss1:\*  
94: gb\_gss2:\*  
95: gb\_gss3:\*  
96: gb\_gss4:\*  
97: em\_gss1:\*  
98: em\_gss2:\*  
99: em\_gss3:\*  
100: em\_gss4:\*  
101: gb\_gss5:\*  
102: gb\_gss6:\*  
103: gb\_gss7:\*  
104: gb\_gss8:\*  
105: gb\_gss9:\*  
106: em\_gss5:\*  
107: em\_gss6:\*  
108: em\_gss7:\*  
109: em\_gss8:\*  
110: em\_gss9:\*  
111: em\_gss10:\*  
112: em\_gss11:\*  
113: gb\_gss10:\*  
114: gb\_gss11:\*  
115: em\_gss12:\*  
116: gb\_gss12:\*







FEATURES	Location/Qualifiers
D6	923 YTYTYTTTTYTTCTCTYYTCTYTCYCYYYYYYYYTCYT 982





DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACRO5N11 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL061936

VERSION AL061936.1 GI:4940214

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr.

- Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/TheBDGP/Drosophila\\_melanogaster\\_BAC\\_library](http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library)

Aaron Mammoler in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the pi and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers

1. .1101

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_lib="RPCI-98"

/clone="BACRO5N11"

/note="end : TET3"

BASE COUNT 631 a 7 c 28 g 289 t 146 others

ORIGIN

Query Match 3.8%; Score 83; DB 122; Length 1101;

Best Local Similarity 40.2%; Pred. No. 6.7e-05;

Matches 210; Conservative 61; Mismatches 245; Indels 6; Gaps 1;

QY 740 tattatttttttcttaaaacaaatttcagcagcagccatgcaaaac 799

DB 1043 TTTTCTTTTNNMAAATATTHMCACCTTTTCATMCTTCWTATTTCTMTM 984

QY 800 ctataaaatattcacacattgaccattgaaagtcgtctcccatgggtaaccagat 859

DB 983 TTTTAAATTTAAATTTATTTCTTACHATTYTTACMYCMWYTTACMYCMWYTTAC 924

QY 860 caaactcacatcccaataacatgagatctctccaccatcatacaattatt----- 914

DB 923 MCAWTTTAAWACTTAAACAAATATTAATTAATTAATTAATTAATTAATTAATTA 864

QY 915 -ttg99gttaaatatcatatttttttagagatttaataagaataataaagattttt 973

DB 863 ATTTAAAAAAWTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 804

QY 974 aaaaaatgataaaattatattcatcatgatttttcatacatattgatttgataataa 1033

DB 803 ATTTAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 744

QY 1034 tataatt 1093

DB 743 TWT 684

QY 1094 tgtttcaaaagcattcatatttaacatttaaaataatttaataactaacagtagaat 1153

Db 683 WATWATAAAWAAWTTWTATTTTWTATTAATAAAWAAWAAWAAWTTTAAAW 624

QY 1154 ctctgtgagtggtggtgagtaggcaacctggcattgaaacgagagagagagtcag 1213

Db 623 TTTTCTTTTAAWTTAAWTTAAWTTAAWTTTATTAATTTTWTATTAATAAA 564

QY 1214 aaccagaagacaataaaatgtagcaacaacaaacaaacaaacaaacaaac 1255

Db 563 TTAWAAAAAATAAAWTTAAWTTAAWTTAAWTTAAWTTAAWTTAAWTTAAWTT 522

RESULT 11

CNS016LI 1101 bp DNA GSS 26-JUL-1999

LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC

DEFINITION BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL106896

VERSION AL106896.1 GI:5624374

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC library (DrosBAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

Location/Qualifiers

1. .1101

/organism="Drosophila melanogaster"

/plasmid="pBelOBAC11"

/db\_xref="taxon:7227"

/clone\_lib="DrosBAC"

/clone="BACN16D22"

/note="end : T7"

BASE COUNT 203 a 220 c 84 g 158 t 436 others

ORIGIN

Query Match 3.8%; Score 82.8; DB 123; Length 1101;

Best Local Similarity 33.6%; Pred. No. 7.2e-05;

Matches 113; Conservative 90; Mismatches 133; Indels 0; Gaps 0;

QY 923 aatattaatcattatttttaagatatattaagaataataaagattttttaaaaaatg 982

Db 1085 AWYWWYTTWYWWYTTCTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTW 1026

QY 983 tataaaattatattcatcagatttttcacacatttgatttgatttgataataatatttt 1042

Db 1025 TAATHTAWTWYCAAAHTAHWWWWTTTATATTAATTAATTAATTAATTAATTAAT 966

QY 1043 ttcaattcttaaaaaatgttgcagacacattattagacatagcttctgtctgttaca 1102

Db 965 WTYTYTATWAWAAWAAWATAASTAAAYWHAATTAATWAWAWATAATTTWTTWTTWTTWA 906

QY 1103 aagcattcatcattatcacattataaaatttttaataactaacagtagaactcttgtg 1162

Db 905 WTAATATWATWAAATTAATWAAWAAWAAWTTAYAWWWAAWTTATWTTATATAATTAAT 846

QY 1163 agtcggtggtgagtaggcaacctggcattgaaacgagagagagagagagagagag 1222



[illegible]

```

QY 972 ttaaaaaaagtataaaattattattcatgattttttcatcacatttgatttgataata 1031
DB 737 TTTTATWTAWATAWAAAAAATAAATAATTTTWTATTTTAAATTTT 678
QY 1032 aatataatttttttcttaaaaaatgttgacagacacttattagacatagcttgg 1091
DB 677 TATTWTTTTWTATATATAAATAWATAWATAATTTATATATATATATATATATAT 618
QY 1092 tctgtttacaaagcatcattcatttaacattacatacaataattttaactaacagtaga 1151
DB 617 WTATTWATATATAAATWTATTTTAAATAWATAWATAAATAAATAAATAAATAA 558
QY 1152 atctcttgt 1161
DB 557 AAATAWTTT 548

RESULT 14
CNS016E1/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN15C13 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL106627.1 GI:5622848
VERSION
AL106627
KEYWORDS
GSS.
ORGANISM
Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1201)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
source
1..1201
/organism="Drosophila melanogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN15C13"
/note="end : SP6"
BASE COUNT 448 a 136 c 161 g 336 t 120 others
ORIGIN

Query Match 3.8%; Score 81.2; DB 123; Length 1201;
Best Local Similarity 44.8%; Pred. No. 0.00013;
Matches 137; Conservative 46; Mismatches 121; Indels 2; Gaps 1;

QY 860 caaacctcacatcaaacataacagatgatactctcttaccatcactactattttggg 919
DB 1190 MMMAWMMWMMHMTCTTTTTTTTATATATAAATAAATAAATAAATAAATAAATAA 1131
QY 920 ttaaatcattacatttttcaagataataaagaataaagaatttttaaaaaa 979
DB 1130 TTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 1071
QY 980 atgtataaaattattattcatgattttttcatcacatttgatttgataataa--atata 1037
DB 1070 ATTTTWTATTTTWTATTTTWTATTTTWTATTTTWTATTTTWTATTTTWTATTTT 1011

```

```

QY 1038 ttttttaattttcttaaaaaatgttgacagacacttattagacatagcttggctgtt 1097
DB 1010 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 951
QY 1098 tacaagaacatcattcatttaacatttaaaaaattttaaactaacagtagaattcttc 1157
DB 950 TTTWAAAAAATAAATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 891
QY 1158 ttgtga 1163
DB 890 TTTTAA 885

RESULT 15
CNS0145P
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN11G11 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL103735.1 GI:5615346
VERSION
AL103735
KEYWORDS
GSS.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1043)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
source
1..1043
/organism="Drosophila melanogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN11G11"
/note="end : T7"
BASE COUNT 277 a 96 c 121 g 382 t 167 others
ORIGIN

Query Match 3.7%; Score 81; DB 123; Length 1043;
Best Local Similarity 41.2%; Pred. No. 0.00015;
Matches 153; Conservative 58; Mismatches 158; Indels 2; Gaps 1;

QY 887 tatctcttaccatcactactaattatttttgggttaataattatcatttttaagat 946
DB 548 TCTTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 607
QY 947 attaataagaataaataaagattttttaaaaaatgtataaataattattatcatgatt 1006
DB 608 AAAAAATATATATATATTTTWTATTTTAAATTTTAAATTTTAAATTTTAAATTTT 667
QY 1007 ttctacattgatttgataaataatttttttcaatttttcaatttttcaaaagtgtgca 1066
DB 668 TTTTWTATATATATATATATATATATATATATATATATATATATATATATATATAT 727
QY 1067 agacacttattagacatgcttggcttcttcacaaagcattcatcatttaacattca 1126
DB 728 ATTTATATATTTTAAATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 787

```

```

QY 1127 aaaaatttt--aatactaacagtagaattcttctgtgagtggtggtgagtaggcaacc 1184
Db 788 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 1185 tggcattgaacgagagagagagagtcagaccagagacaaataaaaaagtatgcaacaa 1244
Db 848 TAWAWAWAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAHAYATACMAAAA 907
QY 1245 acaaatcaaaa 1255
Db 908 MCWAATHAWAA 918

```

Search completed: September 2, 2000, 00:25:01  
 Job time: 4325 sec







Query Match 2.58; Score 53.2; DB 1; Length 3138;











Search completed: September 2, 2000, 01:59:26  
Job time: 8330 sec

[illegible]

CC encountered the problem where if an adenovirus vector capable of  
CC expressing an apoptosis-associated gene is introduced into animal cells,  
CC the cells producing the virus will be destroyed because the period of  
CC time required to induce cell death by apoptosis is shorter than that  
CC required to replicate and produce the virus, resulting in failure to  
CC obtain a recombinant virus having the integrated apoptosis-associated  
CC gene. In this invention an apoptosis-resistant 293 cell line (having an  
CC apoptosis resistant gene introduced) is established and overcomes the  
CC problem. The present sequence represents the base sequence of the  
CC plasmid pRc-Bcl-xl-bsr, which contains the human Bcl-xl gene, and  
CC is used in an example from the present invention.  
SQ Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T;

Query Match 3.0%; Score 65.8; DB 1; Length 7372;  
Best Local Similarity 43.7%; Pred. No. 0.0083;  
Matches 289; Conservative 0; Mismatches 372; Indels 0; Gaps 0;

Qy 1491 agttgagaccagacacactcgttcacatatactctctctctctctctctctctctctctc 1550  
Db 5170 AGCTTCGACCAATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5111  
Qy 1551 tcagggtactttctctccctcaccacaaatccatagattccggtggttcaattcggatct 1610  
Db 5110 TTTTCTT 5051  
Qy 1611 tgcactctggttctgcttgcctcttctcccaactcgggtccatctaggatccatctg 1670  
Db 5050 TTTTCTT 4991  
Qy 1671 gaaactctactctctcttataatctcgcggaatacgcgttggtgactctcagatctcga 1730  
Db 4990 TTTTCTT 4931  
Qy 1731 aatcatttcataaagcct 1790  
Db 4930 TTTTCTT 4871  
Qy 1791 attcacaataaaccttgggctctgctgactgagatggggttctggtgattacagaatt 1850  
Db 4870 TTTTCTT 4811  
Qy 1851 ttacgaaatttgtaattgtaattgtaattgtaattgtaattgtaattgtaattgtaattg 1910  
Db 4810 TTTTCTT 4751  
Qy 1911 atacattctctaggtctcaattttatttcgagatagtcacaaataggaaattcaaaccttg 1970  
Db 4750 TTTTCTT 4691  
Qy 1971 agcagggaataataccctctcctccaaatccagtttctgtgtatataatgttttaaaaatg 2030  
Db 4690 TTTTCTT 4631  
Qy 2031 aaacttttgccttataatctctataataacttttttcttgccaaaatttttgcagtgctct 2090  
Db 4630 TTTTCTT 4571  
Qy 2091 ttgctctccctgtgtataaattctactgttaggtactactactactactactactactct 2150  
Db 4570 TTTTCTT 4511  
Qy 2151 t 2151  
Db 4510 T 4510

RESULT 3  
X33180/c  
ID X33180 standard; DNA; 7797 BP.  
AC X33180;  
DT 23-JUN-1999 (first entry)  
DE Cowpox virus bar full length gene sequence.

Qy 1611 tgcactctggttctgcttgcctctctctctcaactcgggtccatctaggatccatctg 1670  
Db 4322 TTTTCTT 4263  
Qy 1671 gaaactctactctcttataatctcgcggaatacgcgttggtgactctcagatctcga 1730  
Db 4262 TTTTCTT 4203  
Qy 1731 aatcatttcataaagcct 1790  
Db 4202 TTTTCTT 4143  
Qy 1791 attcacaataaaccttgggctctgctgactgagatggggttctggtgattacagaatt 1850  
Db 4142 TTTTCTT 4083  
Qy 1851 ttacgaaatttgtaattgtaattgtaattgtaattgtaattgtaattgtaattgtaattg 1910  
Db 4082 TTTTCTT 4023  
Qy 1911 atacattctctaggtctcaattttatttcgagataggtcacaataggaaattcaaacctttg 1970  
Db 4022 TTTTCTT 3963  
Qy 1971 agcagggaataataccctctcctcacaactcagtttctgtgtatataatgttttaaaaatg 2030  
Db 3962 TTTTCTT 3903  
Qy 2031 aaacttttgccttataatct 2090  
Db 3902 TTTTCTT 3843  
Qy 2091 ttgctctctggttctgtaatttactgttctaggtactactactactactactactactct 2150  
Db 3842 TTTTCTT 3783  
Qy 2151 t 2151  
Db 3782 T 3782

RESULT 2  
X33182/c  
ID X33182 standard; DNA; 7372 BP.  
AC X33182;  
DT 25-JUN-1999 (first entry)  
DE Base sequence of the plasmid pRc-Bcl-xl-bsr.  
KW Cowpox virus; bsr: viral vector; expression; apoptosis; resistance;  
KW crmA; bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;  
KW autoimmune disease; graft rejection reaction; inflammation;  
KW inflammatory disease; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN W09913073-A2.  
PD 18-MAR-1999.  
PF 07-SEP-1998; J04010.  
PR 08-SEP-1997; JP-259235.  
PA (RPRG-) RPR GENCELL-ASIA PACIFIC INC.  
PI Hamada H.  
DR WPI; 99-243728/20.  
PT New apoptosis-resistant virus-sensitive cell  
PS Example 2; Page 41-45; 51pp; English.  
CC The present invention describes an apoptosis-resistant virus-sensitive  
CC cell line into which an apoptosis resistance gene has been introduced.  
CC The recombinant viruses generated are capable of expressing apoptosis-  
CC associated genes. These can then be used in a variety of diseases for  
CC which the induction of apoptosis by gene transfer, or where the  
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses  
CC are useful as vectors for gene therapy which can be applied to cancer  
CC therapy for destroying cancer cells selectively, the treatment of  
CC autoimmune diseases and graft rejection reaction, and apoptosis induction  
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have













CC the detection, diagnosis, characterisation, prevention and therapy of  
CC Bb infections, e.g. Lyme disease. They can also be used for the  
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs  
CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
CC Lyme disease.  
SQ Sequence 116277 BP; 42656 A; 19868 C; 14490 G; 39250 T;

Query Match 2.7%; Score 58; DB 1; Length 116277;  
Best Local Similarity 48.8%; Pred. No. 0.15; Mismatches 157; Conservative 0; Indels 0; Gaps 0;  
Matches 157; Conservative 0; Mismatches 157; Indels 0; Gaps 0;  
QY 65 tatctgtttatgatcttcaggcgcaaaatgcgagctacttaataaaattttacattt 124  
DB 18742 TCTCTTCTGTTTCAATTTGCTTATAAATTTTGTAACTTTCAGAAAAATTTAAATTT 18683  
QY 125 aaattagaattttttatcaataaataattattatttagttttttattgagaataattatt 184  
DB 18682 TGATTTTCTTCTTCATTTAAATATATCCATGGATTTTGTTCATATTAATAAGA 18623  
QY 185 agaaaaatttgatcccgattctccctctttcttcgctattcatctttcttaacca 244  
DB 18622 CTTCTGTTTCAATTTTAACTTTTAAAGTGTCTACAAAAATAAATTTATTTA 18563  
QY 245 aaccaattcttatgttcttcaaaatagaacttgaaattatttaataataaaactgaa 304  
DB 18562 TTGTAACTTACTTTTAAATTAATGATTAATAATTAAGGAGAAATTTTATGTA 18503  
QY 305 acaatttggtatcaatcaatacatcgttagtaataaaatgcgataattattgataa 364  
DB 18502 TAAATATGTTTTTTTAAAACTATTGTTCATTTTAAATTTTGTAGTAATGCTTG 18443  
QY 365 atctgcaaaagattttacaatt 386  
DB 18442 TACTTCAAAAGATAGCTCAAT 18421

RESULT 12  
Q03875  
ID Q03875 standard; DNA; 3095 BP.  
AC Q03875;  
DT 24-AUG-1990 (first entry)  
DE Sequence encoding carboxylic terminal part of native GLURP.  
KW Plasmodium falciparum; antigen; malaria; vaccine; GLURP; ss.  
OS Plasmodium falciparum.  
FH Key Location/Qualifiers  
FT cds 1..2352  
FT /tag= a  
FT /product=GLURP  
PN W09022811-A.  
PD 22-MAR-1990.  
PF 18-SEP-1989; 00218.  
PR 03-MAR-1989; US-218985.  
PR 03-MAR-1989; DK-005191.  
PA (STAT-) Statens Seruminst.  
PI DZiegler M, Borre M, Jepsen S, Vuust J, Rieneck K, Wind A, Jakobsen PH;  
DR P-PSDB: R05804.  
PT Polypeptide(s) derived from Plasmodium falciparum antigen - used in  
PT vaccines and in production of antibodies, for diagnosis and  
PT therapy of malaria.  
PS Disclosure; Fig 7; 108pp; English.  
CC An open reading frame of 2349 bps extends from the 5' terminal end of the  
CC sequence displays some of the characteristics of other malaria nucleic  
CC acid sequences: tandemly repeated motifs, high AT content and a  
CC corresponding preference for codons containing these bases, and a high  
CC content of codons for glutamate. Three major repetitive sequences are:  
CC one motif from bp 34 to bp 156 is repeated from bp 289 to bp 411; another  
CC motif from bp 477 to bp 521 is repeated tandemly twice from bp 522 to bp  
CC 566 and from bp 567 to bp 611; a third motif from bp 1174 to bp 1233 is

CC repeated tandemly 11 times. This last repetitive region consists of 360bp  
CC repeats differing only in 3 bases GAT coding for aspartate. This region  
CC is flanked to the 5' terminal of a degenerated 60 bp repeat. GC content  
CC of the coding part of the insert is on average 30%, and of the non-coding  
CC 3' terminal 11%.  
SQ Sequence 3095 BP; 1443 A; 300 C; 491 G; 861 T;

Query Match 2.6%; Score 57.2; DB 1; Length 3095;  
Best Local Similarity 50.3%; Pred. No. 0.19; Mismatches 167; Conservative 0; Mismatches 163; Indels 2; Gaps 1;  
QY 101 agtacttaataaaattttacatttaaaattagaattttttttatcaataaaattatttta 160  
DB 2225 AGAAATTTTAAAAAGATATCTCAAACTATTGTAAGTGTATGATTAATGCATATGATGTG 2284  
QY 161 ttatttttttagaataatttaattagaataatttgaatccccgattttctctctctttct 220  
DB 2285 TTATTCAGTTGTAGTACATTAAGGAATAGCAAGATATAGTAATATT--TTTCCA 2342  
QY 221 tegtatttcattctttcttaacaaacaaacttttatgtttcttcaaataggaaacttgaa 280  
DB 2343 AAACATTTTAAATAATTAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 2402  
QY 281 attattattataatttaactgaacaaacttttggtatcaatcatcatatcatatcgtcttagtaa 340  
DB 2403 ATTTTCTTCTTATATATGTAACATTAATTTTATTAATAAATAAATAAATAAATAAATAA 2462  
QY 341 laaaatgcgataaatttaattgataaactctgcaaaagattttcaaaatctcttcagaaaaa 400  
DB 2463 AAGAATACTAGGATTTGTTGATATATAGAAATAAATAAATAAATAAATAAATAAATAA 2522  
QY 401 attataaacaatttttgcgtttttcatggtgt 432  
DB 2523 AATAAAATAAATAATTTCTTTCTTAGTTGT 2554

RESULT 13  
Q03568  
ID Q03568 standard; DNA; 6124 BP.  
AC Q03568;  
DT 02-AUG-1990 (first entry)  
DE Sequence encoding the SERA protein of Plasmodium.  
KW Malaria; SERA protein; vaccine; ds.  
OS Plasmodium falciparum.  
FH Key Location/Qualifiers  
FT misc\_feature 1..6  
FT /tag= a  
FT /note="Sequence obscured."  
FT misc\_feature 121..128  
FT /tag= b  
FT /note="Sequence obscured."  
FT misc\_feature 241..245  
FT /tag= c  
FT /note="Sequence obscured."  
FT exon 2407..2440  
FT /tag= d  
FT intron 2441..2598  
FT /tag= e  
FT exon 2599..3405  
FT /tag= f  
FT intron 3406..3580  
FT /tag= g  
FT exon 3581..3721  
FT /tag= h  
FT intron 3722..3850  
FT /tag= i  
FT exon 3851..5835  
FT /tag= j  
FT allele 3087  
FT /tag= k  
FT allele 3092  
FT /label=Allele.II.





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 02:13:13 ; Search time 5217.98 Seconds  
(without alignments)  
740.491 Million cell updates/sec

Title: US-09-464-528-14  
Perfect score: 2165  
Sequence: 1 atcgatagacatgtatt.....gttttgaagtataaagatg 2165

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0

Maximum Match 100%

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

GenEmbl.\*

1: gb\_bal.\*

2: gb\_ba2.\*

3: gb\_om.\*

4: gb\_ov.\*

5: gb\_pat.\*

6: gb\_ph.\*

7: gb\_pl1.\*

8: gb\_pl2.\*

9: gb\_pr1.\*

10: gb\_pr2.\*

11: gb\_pr3.\*

12: gb\_ro.\*

13: gb\_sts.\*

14: gb\_sy.\*

15: gb\_un.\*

16: em\_fun.\*

17: em\_hum1.\*

18: em\_hum2.\*

19: em\_in.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_sy.\*

29: em\_un.\*

30: em\_vl.\*

31: gb\_htg1.\*

32: gb\_htg2.\*

33: gb\_in1.\*

34: gb\_in2.\*

35: em\_bal.\*

36: em\_ba2.\*

37: em\_hum3.\*

38: em\_hum4.\*

39: gb\_pr4.\*

40: gb\_htg3.\*

41: gb\_htg4.\*

42: gb\_htg5.\*

43: gb\_htg6.\*

44: gb\_htg7.\*  
45: em\_htg1.\*  
46: em\_htg2.\*  
47: em\_htg3.\*  
48: em\_hum5.\*  
49: gb\_pl3.\*  
50: gb\_pr5.\*  
51: gb\_htg8.\*  
52: gb\_htg9.\*  
53: gb\_htg10.\*  
54: gb\_htg11.\*  
55: gb\_htg12.\*  
56: gb\_htg13.\*  
57: gb\_htg14.\*  
58: gb\_in3.\*  
59: gb\_htg15.\*  
60: gb\_htg16.\*  
61: gb\_htg17.\*  
62: em\_htg4.\*  
63: em\_htg5.\*  
64: em\_htg6.\*  
65: em\_htg7.\*  
66: em\_hum6.\*  
67: gb\_htg18.\*  
68: gb\_htg19.\*  
69: gb\_htg20.\*  
70: gb\_htg21.\*  
71: gb\_htg22.\*  
72: gb\_htg23.\*  
73: gb\_htg24.\*  
74: gb\_htg25.\*  
75: gb\_htg26.\*  
76: gb\_htg27.\*  
77: gb\_htg28.\*  
78: gb\_htg29.\*  
79: gb\_htg30.\*  
80: gb\_htg31.\*  
81: gb\_vil.\*  
82: gb\_vil2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	118.4	5.5	4601	33	DMU11584	U11584 Drosophila
2	118.4	5.5	19517	58	DMU37541	U37541 Drosophila
3	99.8	4.6	86829	33	PFMAL3P5	AL034556 Plasmodi
4	98.4	4.5	86829	33	PFMAL3P5	AL034556 Plasmodi
5	97.4	4.5	14867	34	AE001398	AE001398 Plasmodi
6	92.6	4.3	146285	39	AC005083	AC005083 Homo sapi
7	91.8	4.2	67970	33	PFMAL1P3	AL031746 Plasmodi
8	91.2	4.2	153098	33	PFMAL3P2	AL034558 Plasmodi
9	91.2	4.2	5371	7	YSCMTTGA	J01462 Yeast (S.ce
10	89.6	4.1	67970	33	PFMAL1P3	AL031746 Plasmodi
11	89.6	4.1	104992	41	AC003504	AC003504 Plasmodi
12	89.6	4.1	130281	60	AC004157	AC004157 Plasmodi
13	89	4.1	161891	55	AC008206	AC008206 Drosophil
14	88.8	4.1	253307	33	PFMAL3P7	AL034559 Plasmodi
15	88.2	4.1	80518	31	PFMAL13PA	AL109815 Plasmodi
16	88	4.1	170125	41	AC007465	AC007465 Homo sapi
17	87.4	4.0	4601	33	DMU11584	U11584 Drosophila
18	87.4	4.0	19517	58	DMU37541	U37541 Drosophila
19	86.8	4.0	104992	41	AC005504	AC005504 Plasmodi
20	86.8	4.0	130281	60	AC004157	AC004157 Plasmodi
21	86.8	4.0	152209	11	HS1108D11	AL034419 Human DNA
22	86	4.0	15421	33	PFCOMP1RA	X95275 P.falciparu
23	85.2	3.9	14001	33	PFCOMP1RB	X95276 P.falciparu
24	85.2	3.9	15421	33	PFCOMP1RA	X95275 P.falciparu

















Qy	781	ctcgcccatgaaaaaaccttataaaadattcctacacattgaccattgaaaagtctggtt	840
Db	40232	TTTAAATTAAATTAATAATTAATTTATTTATATATATAATTAATTTATATAATATATATTT	40291
Qy	841	ctcccaggggttaaccagatcaaacctcacatccaacaacataacatgatctccctaccacaa	900
Db	40292	ATTAAATTTTATTTTATTTTATTT	40351
Qy	901	tcatactaattatttttggtttaaatatttaattcattatttttaaagatatattaataagaatt	960
Db	40352	TATTATTTTAAATATATATTTAAATTTAAATTTATTTTATATTTTAACTTTAATTAATTAAT	40411
Qy	961	taaaagatt--ttttaaaaaaatgataaaaatttatattattcatgattttttcatcacatttt	1018
Db	40412	ATTAATATTAATAATTAATAATTTATTTAAATAATAATAATTTTAAATAATAATTTAAATTT	40471
Qy	1019	gatttgtataaataatatatttttttcaaattcttcaaaaagtgtgcgaagacacttatta	1078
Db	40472	AATTATATATACATTATTAAGCATTTTTATTTAATTAATATATTTATTTAAATTTAATTTATTT	40531
Qy	1079	gacatagctctgttctgttttcaaaaagcattcatcattttaacattaaaaaattattaa	1138
Db	40532	ATATTAAAGATTTTATTTTAAATTAATATATTTATTTAAATTTAAATTTAATTTATTTAT	40591
Qy	1139	tactaacagt	1148
Db	40592	TATATATAAT	40601

RESULT 5  
 AE001398 LOCUS  
 DEFINITION Plasmodium falciparum chromosome 2, section 35 of 73 of the complete sequence.  
 ACCESSION AE001398 AE001362  
 VERSION AE001398.1 GI:3845197  
 KEYWORDS malaria parasite P. falciparum.  
 SOURCE Plasmodium falciparum  
 ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 14867)  
 Gardner,M.J., Frettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L., Koonin,E.V., Shallow,S., Mason,T., Yu,K., Fujii,C., Pederson,J., Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Perlea,M., Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O., Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L. Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum Science 282 (5391), 1126-1132 (1998)

JOURNAL MEDLINE Erratum:[published erratum appears in Science 1998 Dec 99021743  
 REMARK 4:282(5395);1827]]  
 REFERENCE 2 (bases 1 to 14867)  
 Gardner,M.J.  
 Direct Submission  
 TITLE Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA

FEATURES source  
 location Qualifiers  
 1..14867  
 /organism="Plasmodium falciparum"  
 /db\_xref="taxon:5833"  
 /chromosome="2"  
 complement(1570..2424)  
 /gene="PF0490C"  
 complement(1570..2424)  
 /gene="PF0490C"  
 /note="predicted by GlimmerM"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAC71887.1"  
 /db\_xref="GI:3845198"  
 /translation="MKENKIKIMDYLSCLPDDVVREKSGKNLSLKSSSTKKSDYK"  
 /translation="MKENKIKIMDYLSCLPDDVVREKSGFKPKYYNNRLINKNYNNGRNFHNG"  
 SSIFSFKROSHKHGSGFRGRSGFIRNKSGFPKPKYYNNRLINKNYNNGRNFHNG"

DNFKGRTGSGSVFDRNRKSGFKKRFRTISNRNKSVSVKYGRONGSNKMGKRSFNKAPTST  
TVYTKRLNNYKTVSAPYKPKFNNNLSILYRKNRFTFALNTKSKPGVTIKSSVPKRRKIKLSL  
GLKGSLSKSTRKSTSGSKFKPLNKYFLSKIKIVTSLNKKIPSLPKQKNTFVALPESL  
NNATTKN"  
11241..14606  
/gene="PF80495w"  
11241..14606  
/gene="PF80495w"  
/notes="predicted by GlimmerM"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAC71888.1"  
/db\_xref="GI:3845193"  
/translation="MFALKNTVREGFVNICFSLKLYLKSNFVTNLYNRYENNRKRR  
TNKIKYKSAQSLDGLNIHOKLILFKMLPKYKCAKCYISAKERYVTKLLEIDYKCK  
FNTISGDIQTSYKIFDELDTKTDFYIEVANKIDNKNLKNIEYIFKNDKIDFHRRL  
EYLIGCKIKMSIYKHEMNGNELTHFLYIFRWNRKNDKNLILFYNYFYFDHMLFNL  
HFYKLLFINKYLNKNSNIPFNKNLQIOMEFNLYPERIKNEKNYIKNKNKEIYKKK  
CPAKFHEVDHIDNEKILNLRILYVDNSILOIDINNKMLNANNLNINTEYISKLL  
NFVCTLLKKGKYDNMTYIKLKEVIKATHILCDKTKNLETFCSDIDYSTLLASLNKK  
FLNKIDDLKFLILNLINKEVNFQSCISILKNIYNNINLRNNYIVTVNNV  
LNDIMKFSYLNIFLGKRIKTEENAVLIIHNDDQTNSENKNIKDIIOIKRKEY  
IRFYMENYKOHFKRLKSDLSLKLSTFVKINEVNSDFYLLFNFIICILYNFVL  
NRNSVKYKOTYIYLNDSLFVYKIKNDRTKKKNFFLLSSMKELICKNLSVSN  
RYIKLHEEDNFQDKDOYCSLFLNLFDPKIIHFHYIYNLACHVYKTYNPKCNKL  
INDIISLLTSCSKPQFIENNSNDQSKRELILKYNIIIDILKNTLNYKISID  
NKGIFLSLSNYSKTCVENLNLDSLOSEFEKVTKSKGGIHMNDLNNNNCS  
KPHRYEYKEKNLFINLKIIECLINHLFKLKKTYLYLKQSLCPINKLENIL  
KKIYIANNLYEMVYGVCEMLERVLSSKQONLFYSYNNKVNKFMKOFILCHISE  
DXYIEMSTNVFLFYDLKLNINSEROSNILRNNSNDRFIDEIKKYKLNNTNLKHK  
NNVYLNKESNNSNGNLSNLKODKKNHNHNVENDLIDKNNEKKIOEKQNGCENEC  
CKOVLVNDIINIFGLAMEKKELFFPOLYMLCNITAFKRYIVSSSLFPHMDVFKIILK  
DMLKYLCLENYKIKNECAFLYITDVLFKER"

BASE COUNT 6284 a 1019 c 1106 g 6458 t  
ORIGIN

	Query Match	4.58;	Score 97.4;	DB 34;	Length 14867;
	Best Local Similarity	46.8;	Pred. No. 0.00045;		
	Matches 483;	Conservative 0;	Mismatches 531;	Indels 17;	Gaps 5;
Qy	107	taataaattttacattaaattagaattttttttatcaataataattcaattatttaggt	166		
Db	7390	TAATAATAATATTTAATAAACAATAATTAAATAT-TTAATAATAATTATATATAATAT	7448		
Qy	167	ttattagaataattaaattagaaaaattttgaacccccgatttctctctcttcctcgcta	226		
Db	7449	ATATATAATAATTATTATTATTATTATTATAAAAAATATATATAATAATAATAATTA	7508		
Qy	227	ttcatcattttctaaacaaaccaatttatatgctttcctcaattagaaactgaaattatt	286		
Db	7509	TTACACGTGTTATTTATATATTAATTAATTTATTTTAAATATATAATAATATATATATTA	7568		
Qy	287	aattataattaaactgaaacaaatttggatccaattccaattacatcgcttagtaaaaaaat	346		
Db	7569	ATTTAAATAAACAATAATATATGTTTATTAATTTTAATAAATAAATAATATATATATT	7628		
Qy	347	gcgataaattaatgataaactgcgaagaattttcacaaatactcttcgagaaaaaattaat	406		
Db	7629	ACATATATAATTAACTT-----AAAAATAAATAAATAAATAAATAAATTTCATTTATTACC	7682		
Qy	407	aacaaattttctgcttttcagggtgtggtctgaggaggattttggcactatagaactctc	466		
Db	7683	ATTTAAATTAAATAATAAATAAATAATATTTTAAATTTAAAAATTAATTAATAATTAATA	7742		
Qy	467	ctacggaccattcttgcactccaactaaacgatggctcagaattgggtgggattttatat	526		
Db	7743	AAATATATATATTATATAAATTTAATATTAAAAATAAATTTAATTAATAAATAAAT	7802		
Qy	527	tcaagcatatccctttcaaaacttccctacttactcgctcgcttgcgtaactcggtaacatt	586		
Db	7803	AAATAAATAAATAAATAATATATATATTTTAAATTAATCTATTTGTTATCTATAATAAT	7862		



```

QY 587 agacttccaaatcatttttaacccttaaacagtaaatgtgaaggacaaaataattt 646
D 7863 AAAACTATTATA-----TTTAAATATATAAAATCGTATATAATAATAATCTTTTAA 7918
QY 647 ttcaaatgtgacagactatttttttttttttttttttttttttttttttttttt 706
D 7919 TTTTATTATTATAATTAATTAATTTATTATTATTATTATTATTATTATTATTATTAT 7978
QY 707 gaatttttagaacccagatgttaactaaacccaatttttttttttttttttttttt 766
D 7979 TTTTAAATGTTTAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 8038
QY 767 tcatgagcagatgcctccagccctgaataaaacccctataaaaatactcacacattgacca 826
D 8039 ATAATTATAATTATAATATATATTTTACATATAATAATAATAATA-----TTATATATATTTAA 8093
QY 827 ttgaaagtgcttcctccatcggttaaacacagatcaaacctcacacatacaacatgg- 885
D 8094 TTTTATTATTTTTATTAAATGAATTAATAATAATAATAATAATAATAATAATAATAATT 8153
QY 886 atactccttaccacatacactaattatttgggttcaaatatttaatttttttaaga 945
D 8154 AAATATTATTATATATATTTAAATTAATAATAATAATAATAATAATAATAATAATAATA 8213
QY 946 tattaattaagaataataaagatttttttaaaaaaaagtataaaattatatttcgat 1005
D 8214 ATTAATATTATTATTATAAAAAAATAATTAATAATAATAATAATAATAATAATAATAATA 8273
QY 1006 ttttcacattgttgatataataatttttttttttttttttttttttttttttttttt 1065
D 8274 TATTATTAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8333
QY 1066 aagacacttattagacatagcttctgttttttttttttttttttttttttttttttt 1125
D 8334 ATATTATTATAGTCTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 8393
QY 1126 aaaaaatattt 1136
D 8394 TAATAATATAT 8404

RESULT 6
AC005083/c
LOCUS AC005083 146285 bp DNA PRI 21-DEC-1999
DEFINITION Homo sapiens BAC clone CTA-281G5 from 7p15-p21, complete sequence.
ACCESSION AC005083
VERSION AC005083.1 GI:4150930
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 146285)
AUTHORS Madsen,C. and Blair,T.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 146285)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 146285)
AUTHORS Waterston,R.
JOURNAL Direct Submission
Submitted (12-JAN-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 146285)
AUTHORS Waterston,R.
JOURNAL Direct Submission
Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 12, 1999 this sequence version replaced gi:3212908.

```

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
----- Summary Statistics  
-----  
Center project name: H\_RG281G05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double-stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgrl.nih.gov/DIR/FTP/CHR7>, send [mailto:egreen@nhgrl.nih.gov](mailto:mailto:egreen@nhgrl.nih.gov), or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

Clone CTA-281G5 is from a release of the human BAC library CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).  
VECTOR: pBeloBAC11  
Selection: chloramphenicol

#### NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of CTA-281G5;  
actual end is at 146285 of CTA-281G5

BAC CTA-281G5 contains an E. coli transposon from 10982 to 12345 that is not represented in the submitted sequence.

#### FEATURES

	source
repeat_region	1..146285 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7p15-p21" /clone="CTA-281G5" /clone_lib="CITB-HS-A"
repeat_region	322..358 /rpt_family="TAAAAn" 808..952 /rpt_family="MER1_type"
repeat_region	1172..1194 /rpt_family="AT_rich"
repeat_region	1264..1720 /rpt_family="L1"
repeat_region	1848..1881 /rpt_family="AT_rich"
repeat_region	1882..2171 /rpt_family="Alu"
repeat_region	3782..3804 /rpt_family="AT_rich"
repeat_region	4794..4826 /rpt_family="AT_rich"
repeat_region	5764..6034 /rpt_family="purine-rich"
repeat_region	6891..7126



```
repeat_region /rpt_family="L1"
7133..7263
repeat_region /rpt_family="L1"
7455..7481
repeat_region /rpt_family="AT_rich"
7504..7554
repeat_region /rpt_family="AT_rich"
8320..8348
repeat_region /rpt_family="(CA)n"
8466..8533
repeat_region /rpt_family="AT_rich"
9611..9733
repeat_region /rpt_family="MIR"
9946..10172
repeat_region /rpt_family="MER1_type"
10437..10477
repeat_region /rpt_family="(CA)n"
10581..10681
repeat_region /rpt_family="(TAAA)n"
10759..11049
repeat_region /rpt_family="AT_rich"
11050..11235
repeat_region /rpt_family="(CAATA)n"
11408..11526
repeat_region /rpt_family="(CATAT)n"
11768..11872
repeat_region /rpt_family="(CATTA)n"
12002..12110
repeat_region /rpt_family="(CATAT)n"
12057..12168
repeat_region /rpt_family="(TA)n"
12381..12490
repeat_region /rpt_family="(CAATA)n"
12562..12680
repeat_region /rpt_family="(CTATA)n"
13131..13367
repeat_region /rpt_family="L2"
13378..13540
repeat_region /rpt_family="MER1_type"
13557..13744
repeat_region /rpt_family="L2"
13745..14592
repeat_region /rpt_family="L1"
14593..14616
repeat_region /rpt_family="L2"
16078..16548
repeat_region /rpt_family="L1"
18492..18692
repeat_region /rpt_family="MER1_type"
19334..19422
repeat_region /rpt_family="LFR41"
19880..20137
repeat_region /rpt_family="LFR41"
20504..20541
repeat_region /rpt_family="MER1_type"
21092..21607
repeat_region /rpt_family="MaLR"
22221..22499
repeat_region /rpt_family="Alu"
22500..22523
repeat_region /rpt_family="(CAAAA)n"
22535..22857
repeat_region /rpt_family="MER4-group?"
23203..23380
repeat_region /rpt_family="L1"
23381..23643
repeat_region /rpt_family="Alu"
23644..24770
repeat_region /rpt_family="L1"
25826..25846
repeat_region /rpt_family="AT_rich"
26130..26254
repeat_region /rpt_family="MIR"
```

```
repeat_region 26529..26563
/rpt_family="AT_rich"
repeat_region 26792..27116
/rpt_family="Alu"
repeat_region 30744..30982
/rpt_family="MIR"
repeat_region 31032..31739
/rpt_family="L1"
repeat_region 32332..32448
/rpt_family="MIR"
repeat_region 32536..32730
/rpt_family="MIR"
repeat_region 33282..33333
/rpt_family="AT_rich"
repeat_region 35261..35288
/rpt_family="(CATTT)n"
repeat_region 35571..35598
/rpt_family="(CA)n"
repeat_region 35924..36007
/rpt_family="L2"
repeat_region 36476..37873
/rpt_family="L1"
repeat_region 38324..38899
/rpt_family="Retroviral"

Query Match 4.3%; Score 92.6; DB 39; Length 146285;
Best Local Similarity 44.5%; Pred. No. 0.00089;
Matches: 467; Conservative 0; Mismatches 569; Indels 13; Gaps 2;

QY 109 ataaattttacatttaaaatttagaattttttttatcaataaataattattattattttt 168
DB 11581 ATATAATATATATTTATATATATATATATATATATATATATATATATATATATATAT 11522
QY 169 attagaaattattagaaattttgaattcccgatttctctctctctctctctctctct 228
DB 11521 ATATGTAATATATATTTTCATATATATATATATATATATATATATATATATATATAT 11462
QY 229 catcatttttcaacaaaccaattcttatgttcttcaaaattagaaacttgaaattattaa 288
DB 11461 ATTAATATGATATATATATATATATATATATATATATATATATATATATATATATAT 11402
QY 289 ttataattaaactgaaacaaatttggtatcaattcatcacatgcttagtaataaaatgc 348
DB 11401 ATTATTCATATATCTAAATATATATATATATATATATATATATATATATATATATATAT 11342
QY 349 gataaattatgataaactcgcaaaagattttacaaatattcttccacaaaaaataaataa 408
DB 11341 TATATATATTTATATATATATATATATATATATATATATATATATATATATATATATAT 11282
QY 409 caaattttgtcttttcattggtggtggtggtggtggtggtggtggtggtggtggtggt 468
DB 11281 ATCATTTATATATTTCTATATATATATATATATATATATATATATATATATATATATAT 11222
QY 469 acggaccattctttgcacttcaactaaacgattggtcgaattggtggtggtggtggtggtggt 528
DB 11221 TATAAATATATCTTTTATACATAAATATATATATATATATATATATATATATATATATATAT 11162
QY 529 aagcatacccttttcaaaacttccactactctcgctcgctcgctcgctcgctcgctcgctcg 588
DB 11161 TAATATTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 11102
QY 589 actttcaaaatcatttttaacccctcaaacagtaaatgtgaaggacacaaataatattttt 648
DB 11101 ATTATATACAAATATATATATATATATATATATATATATATATATATATATATATATAT 11050
QY 649 caaattgttagactatttttttttttttttttttttttttttttttttttttttttttttt 708
DB 11049 TTTATATATATACAAATATATATATATATATATATATATATATATATATATATATATAT 10990
QY 709 attttaggaaccacagatgtaactaaacaaataattttatttttttttttttttttttttttt 768
DB 10989 AATGTATCAATTTATATATATATATATATATATATATATATATATATATATATATATATAT 10930
```







```
/gene="PFC0175w"
/Note="predicted using hexExon; MAL3P2.4 (PFC0175w),
Hypothetical protein, len: 594 aa; Similarity with a
number of hypothetical proteins from model organisms.
C.elegans hypothetical protein (TR:O01963) BLAST Score:
361, sum p(1) = 4.0e-47, 35% identity in 227 aa overlap."
/codon_start=1
/protein_id="CAB38990.1"
/db_xref="GI:4493881"
/translaton="MNPKEQNKNQINQIVTYEYQKKKKVVIPIINDTKNNIINHN
YDYSINDKYNNSTDCVLMNNEQDINLYNYYKKKLLKFOENNLNLYNVEKLDIEKKME
LAILNMLNIDQYDAYHFKAYFNNSDNLLENFNNKKVLTLLKSLHKBEDILN
NNNNNIDPMIKQKQFVHCKQKQKFCIPILFLECDIEDTTLSCGHVYKRECLN
YLTSLHNEDEDDIITKCDLCKNKKIKNNDKNIKEEDKSKYLYLTHIVIKKSK
DLKCPNPKPIQSVLNNNNVICKGTHFCFECUHEFRPLCSYIKKWELENN
DNNKWIHAYTKPCNKNPIEKNSCMNVKICGYSFWRCLDNKNHKGFGYKCN
KYLEHNSKYNEQKQKKTKKDDTKYDDEKEDTKDNDNDNQNNEERKNSHL
ILNRYNHFKSRFNQHOVAENFSTHTQLFLFNCKNYNIHLKMKFPEDAIQIKCR
KILKWSYIAYFSNKSNDQKHLFEYHGELEKLNLDLQTKEDINLTQFNKNTDNT
VRDIQITQIMIDIFPKNICEFMENNIV"
/complement(21152..24194)
/gene="PFC0180c"
/complement(join(21152..23273,23390..23604,23713..23862,
23997..24194))
/gene="PFC0180c"
/Note="predicted using hexExon; MAL3P2.5 (PFC0180c),
Hypothetical protein, len: 895 aa"
/codon_start=1
/protein_id="CAB38992.1"
/db_xref="GI:4493883"
/db_xref="SPTREMBL:O97228"
/translaton="MFNACQLKSCNCEAEEMRQSNKGVEDLYGHHIINENSLKK
IIECPDLDAVEIGTEREYVAITVQPDVITRVFVFVETIETTPVKIVYEEK
IKEVPIAVIKVKEVPEVKEVVDKIVDPDIQYCLKYPKVEIQENIKRPVFKK
FVEKIVPVKVKREVEVEYVVIKPKDKSKTKKENETACINETNETNE
EGGRNVERGIYEHPLNLRNGARLIMGPPVQOEGIIIEHTTEMHTSISGTRN
VSCFCNKVQGNROTNCYLSGSCMCPRTMGLKKTNTMLNPRLEOVFPKPKVN
IEVQKHVPSVDVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPV
QVDCIKKPKYKIDQVQKIPCDVDIKYKMEKACAVPNELAQDDVHAMKVRNKH
LAERKKEGEFYFYKSKNHYIYICIAIIFYNFFYFFYFFYFFYFFYFFYFFY
MEENSTNEICEKEEYIYLEDEDERFINKNENINTEIKEDSNGKEINENIKNEI
NENTNKEINENSGKEINENSGKEINENSGKEINENSGKEINENSGKEINENSG
SVTKENHNYKNETNDFALINDEKVIKAEHIFVRNKNIEKAYDIFNYSKEL
RFKDNKMLQNDINLNENKTYLFESDIFENMEKEDIENKNNNTSNKYLEEQ
LTASLYPSHPLAMTKQKIQDTLTKHLYNNDIFRASVNAVNYNQSGNYSVSDIM
RNNEFLKSNFIISFPSPGNIRNENYNNLIQNDLQDNQKSNYEHMNNQINYN
NYNENFENHOIKLVEQIIKEQQQKESKSKCCNYFCOK"
/21152..28729
/gene="PFC0185w"
/21152..28729
/gene="PFC0185w"
/Note="predicted using hexExon; MAL3P2.6 (PFC0185w),
Hypothetical protein, len: 526 aa"
/codon_start=1
/protein_id="CAB38993.1"
/db_xref="GI:4493884"
/translaton="MVKNDNSQKNCNTRNFCNVNDEKKEEGMMSQKDNENYNPND
IHNHNYNYNNYVSEKNEKFSQNSMNOENYFVDCPPONARILKPLIQKIVEIK
PEIEKIEVPOQVYIEKLVPHVILQEKLIHIPKPVHERIKKSKTIFQEKIVE
POIKVIVEVPOVQVQKIEIVPKIMVQERIIIPVKKIVREKIVEIPQIELKNIDI
EKVQIEPIEVQKIDIPYQIVDPVHVEKIVEVPHVQIYRNVISVPPVHPKPV
EIPMAHYTFVEKIVDRNVPVQLVQVFLCPKIEARYKEIPVPHVQRIIEHP
PKDAMNPHLLPLLYQEDNNIEMSTNKNTMEQNKNCIFNLNLRNDSRNSHASPS
VELLHNDKSKQVYNNHTNKKTSIISNPVFNAYNONGTILNQNPNNDINNNHEII
HNNNIPSNNVFKSFHDNVTHNKNINNMPYNNVTLKNYSVNPITPTKTFPANN
IYSGSPVLSSSHTHA"
/complement(29952..31529)
/gene="PFC0190c"
/complement(29952..31529)
/gene="PFC0190c"
/Note="predicted using hexExon; MAL3P2.7 (PFC0190c), PAST
```

```
Query Match 4.28; Score 91.2; DB 33; Length 153098;
Best Local Similarity 45.78; Pred. No. 0.0013;
Matches 432; Conservative 0; Mismatches 508; Indels 6; Gaps 3;
Qy 115 tttacatttaaaatagaattttttttttatcaataaataattatttagttttattaga 174
Db 136984 TTTAACATATAAGCTTTTTTTTTTATACCTACTCTTTTAAATAATAATAATAA 136925
Qy 175 aatatttaattagaataattttgaatcccgagatttctccctcttcttcctgcgtatcatcat 234
Db 136924 TAAATATATATATATATATATATATATATATATATATATATATATATATATAT 136865
Qy 235 tttctaacaacaacaattttatattgtcttcaaataggaaattgaaattataataataa 294
Db 136864 TTATTTCTTTTAATTTATTTTATTTCTTAAATAATTTTAAAAAATATATAATAA 136805
Qy 295 ttaactgaaacaattttgggtatcaattcatatcatgcttagtaataaaatgagataat 354
Db 136806 TTTATATTCACAAAAATATATATATATATATATATATATATATATATATATATAT 136747
Qy 355 taattgataaattctgcaaaagatttttacaataatctttcagaaaaataatacaaaatt 414
Db 136746 ATAATATTAAATATTATATAAAACAAACCTGT-TCCTTTATATATATATAATAA 136688
Qy 415 ttgtcgttttcatggtgtgtgtctgagaggatttggcactatagaactctctctacggac 474
Db 136687 TATTATTTAAACATGGCATATATATATATATATATATATATATATATATATATAT 136628
Qy 475 cattcttgcacttcaactaaacgatggtcagaaattggcgagatttggggattttatatcaagcat 534
Db 136627 TAAATCATATAATTTTATATATTTTATAATTTTATTTTATTTTATTTTATTTT 136568
Qy 535 atccctttcaaaactctactacttactctgcgtcgctgcgtaactcggtaacattagacttcc 594
Db 136567 GAAGTATATAAGTTTTCCTCCATATATAATAATTTATATATAATAATAATAATA 136508
Qy 595 aaaaatcatttttaacccctaacaacagataatttgaaggacaaaaataatttttcaaat 654
Db 136507 TAAATAATATTTATCGGGGATATATAAATTTTATATATATAGACATATATACATTT 136448
Qy 655 tgatagactatttttttttttttttttttttttttttttttttttttttttttttttt 714
Db 136447 CATATACTAATTTATATAATTTTATATAAGCCCTTATAATAAATTTATATATATA 136388
Qy 715 ggaacacacagatgtaactaaacaaatattttatttttcttaaaaaaataatttcattgac 774
Db 136387 TATATAATATATATATATATATATATATATATATATATATATATATATAATAATT 136328
Qy 775 gcatgcctcagccatgaaacacattataaaatatctacacattgaccatttgaaag 834
Db 136327 ACATTCAATAATATATAATAAATTTTCAAAAAAATAATAATAATAATAATAATA 136268
Qy 835 ttcgttctcccatgggttaaccagatcaaacacacacacacacacacacacacacacac 894
Db 136267 TAATATAT-ATTATATATATAATGAATATATATATATATATATATATAAGATATTT 136211
Qy 895 taccatcataactaattattttgggttaaatatttaaatatttttttaagaatttaatta 954
Db 136210 TAAAAATATTTATTCGTATATTTTTTCCAAAAAGAAAAATTTTTTTTTTTTTTAAATATCT 136151
Qy 955 agaaattaaaagatttttttaaaaaatgtataaaattataattattattcatgattttcatac 1014
Db 136150 ATTAATTAATAATATATATATATATATATAATAATAAAAAAACATGGAATTTTTT 136091
Qy 1015 atttgatttgataaataatttttttttttttttttttttttttttttttttttttt 1060
Db 136090 TTTTACTAAAAAATAATATTTCCACTTGTGCTTCTTTATATAT 136045
```

```
RESULT 9
YSCMTTGS/c
LOCUS
```

18-APR-1994

DEFINITION Yeast (S.cerevisiae) mitochondrial Ser-tRNA and ATPase proteolipid genes.

ACCESSION J01462 J01463

VERSION J01462.1 GI:343938

KEYWORDS ATPase; proteolipid; transfer RNA; transfer RNA-Ser.

SOURCE Yeast (S.cerevisiae) mitochondrial DNA; cytoplasmic petite mutants (rho-) ds400/a3 [1], ds401 [2].

ORGANISM Mitochondrion Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1026 to 2806)

AUTHORS Macino,G. and Tzagoloff,A.

TITLE Assembly of the mitochondrial membrane system: The DNA sequence of a mitochondrial ATPase gene in Saccharomyces cerevisiae

JOURNAL J. Biol. Chem. 254, 4617-4623 (1979)

MEDLINE 79173209

REFERENCE 2 (bases 1 to 5371)

AUTHORS Tzagoloff,A., Nobrega,M., Akai,A. and Macino,G.

TITLE Assembly of the mitochondrial membrane system. Organization of yeast mitochondrial DNA in the OLI1 region

JOURNAL Curr. Genet. 2, 149-157 (1980)

COMMENT [2] notes that even though there are other possible coding sequences in the ds401 segment, none are sufficiently long to code for a gene product of the size of the var1 protein. a model for the mitochondrial RNA processing is proposed.

FEATURES

source Location/Qualifiers

1..5371

/organism="Saccharomyces cerevisiae"

/organelle="mitochondrion"

/db\_xref="taxon:4932"

1..5371

/gene="OLI1"

/db\_xref="SGD:S0007274"

1372..1602

/gene="OLI1"

/note="ATPase proteolipid"

/codon\_start=1

/db\_xref="SGD:S0007274"

/transl\_table=3

/protein\_id="AAA32169.1"

/db\_xref="GI:343938"

/translation="MQLVLAAYIGAGISTIGLLGAGIGIAIVFAALINGVSRNPSIKDTVPFMAIFGALSEATGLFCLVLSFLLFGV"

2856..2942

/gene="OLI1"

/note="codon recognized: UCA; Ser-tRNA"

/product="tRNA-Ser"

/db\_xref="SGD:S0007274"

2294 a 352 c 368 g 2191 t 166 others

BASE COUNT 82 bp 5' to SstII site, at wild-type map units 79.1.

ORIGIN

Query Match 4.2%; Score 91; DB 7; Length 5371;

Best Local Similarity 45.7%; Pred. No. 0.0039;

Matches 478; Conservative 0; Mismatches 550; Indels 17; Gaps 4;

QY 16 ttattcaaaacataaataatgatgctcaaaatgggtggtggaacgatactctgttat 75

DB 4287 TTATCTATATCAATTTAACTAATATATATTACTTAAATAATATATATATAT 4228

QY 76 tatgatttcaggcgcaaaaatgcagacttaataaaattttacatttaaatagaatt 135

DB 4227 GATAATTAATAGGTCATAGTACTTTTATTATTAATAATACTAAATATATACAT 4168

QY 136 ttttttcaataaataattattattgatttttagaataatttaattagaataatttg 195

DB 4167 AAATATTATTATATTATTATTATATATTATTATTTTATTATTATATTATATAA 4108

QY 196 aatccccgattctcctctcttcttcgcatactcatcttctcaacaaacccaatcta 255

DB 4107 TAATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATAA 4048

QY 256 tatgttcttcaaatagaacttgaaattattattataattataaactgaaacaaatttgg 315

DB 4047 TATTTAAATTTATTTAAACGTGTTTAAATTAAGTTTACTTATAATATTTTATAATAT 3988

QY 316 atcaattcat--atacatgcttagtaataaaatgggataaataattgataaaatctgcgaa 373

DB 3987 TTATATTATTTAATTTAAAGTTATTTTATATATATTAATTAATTAATTAATTAATCAT 3928

QY 374 agattttacaatatcttttcagaaaaaataataa--caaatgttgcgttttcacatggg 431

DB 3927 TTATAGTATTTTAAATTTCTATATTTTATATATTAATTAATTAATTTCAATTTAT 3868

QY 432 ttgggtcagaggaggttggtgcaactatagaactctcctcagggaccattcttcgacactca 491

DB 3867 TTATATTATTTAAATGTTCTAATTTTATTAATTAATTAATTCCTCCGGGACCAATCCGGTGAA 3808

QY 492 ctaaacgcatggtgcagaatgggtgggattttattatccaagcatatccctttcaaaacttc 551

DB 3807 CAACCGGATGGCGCGCGGTTTATATTAATTAATTAATTTCTTTTATATTTAT 3748

QY 552 ctacttacttcg9cggttcggttaactgtaacttagacttcacaaatcatttttaacccc 611

DB 3747 TTATTTTATTTATTTATTTAT---ATCTAATAATATATTTCTTAATAATTAATTTCTATTTA 3692

QY 612 ctaaacagtaaatgttgaggacacaaataatttttcaaaatttgatagactatttttt 671

DB 3691 TTATGAATCTGATAGATTATTAAATTAATTAATTAATTAATTAATTAATTAATTA 3632

QY 672 ttgttaatttgacgaacacaaa-----ccagatttctcctgaatttttaggaaccac 722

DB 3631 TTGATAAAATTTATTTAATAATTTTAAATTTTCTATTTTATTTTACTTTTATCTA 3572

QY 723 agatgtaactaaacacaaatatttattttctcaacaaacaaatttcacggcagcatgct 782

DB 3571 TTATATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3512

QY 783 cagcccatgaaaaaaccttataaaatctcacattgacattgaaagtcgttct 842

DB 3511 TGCATCCCTTATTAATTTATTTTATTAATTAATTAATTAATTAATTAATTAATTTTG 3452

QY 843 cccatgggttaaccagatcaaacctcacatcccaacataacatggatctccttaccatc 902

DB 3451 CAGTTATTATTAATTAATTAATTTCTTATTAATTAATTAATTAATTAATTAATTAATTA 3392

QY 903 atactcaattatttgggttaaaataatcaatcatttttttaagataatttaagaataa 962

DB 3391 TTATAATTTTATATATGATAATTTATAAAATAAATTAATTAATTAATTAATTAATTA 3332

QY 963 aaagatttttcaaaaaatgataaaattattatttctcattgatttttcacatttgatt 1022

DB 3331 AAATATTATAAATAAATAATATATATTAATTAATTAATTAATTAATTAATTAATTAAT 3272

QY 1023 ttgataataaataatatttttttttaa 1047

DB 3271 ATATATATATATATAAATAATAA 3247

RESULT 10

PFMALIP3 67970 bp DNA -INV 15-DEC-1999

LOCUS Plasmodium falciparum MALIP3, complete sequence.

DEFINITION AL031746

ACCESSION AL031746

VERSION AL031746.9 GI:6594243

KEYWORDS HTG.

SOURCE malaria parasite P. falciparum.

ORGANISM Plasmodium falciparum

REFERENCE 1 (bases 1 to 67970)

AUTHORS Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M. and Barrell,B.

TITLE Direct Submission

JOURNAL Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

## COMMENT

On Dec 16, 1999 this sequence version replaced gi:5763807.  
For more information about this sequence or the Malaria Project,  
see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum). IMPORTANT: This  
sequence is unfinished and does not necessarily represent the  
correct sequence. Work on the sequence is in progress and the  
release of this data is based on the understanding that the  
sequence may change as work continues. The sequence may be  
contaminated with foreign sequence from E.coli, yeast, vector,  
phage etc.

## FEATURES

```

source
    1..67970
        /organism="Plasmodium falciparum"
        /strain="307"
        /db_xref="taxon:5833"
        /chromosome="1"
        complement(1748..3276)
        /gene="MALIP3.01"
        complement(join(1748..2598,2748..2848,2990..3276))
        /gene="MALIP3.01"
        /note="MALIP3.01, conserved hypothetical protein, len: 412
        aa, similarity: UPF0006 family eg to
        YBL055C/YBL0512/YBL0511, YBF5_YEAST (418 aa), fasta
        scores: Opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa
        overlap)"
        /codon_start=1
        /product="conserved hypothetical protein, UPF0006 family"
        /protein_id="CAB63556.1"
        /db_xref="GI:6594244"
        /translation="MKLVFHYIKYINLVFYISIFLKSNSLKIYNDLRYISTVKNKYK
        VOIKRNLKNNHNRHNDNESSFDIGSNLTKDMPGVYNSKKHNDLQNLNRAK
        NNVDKIIITCTCAEIDKSLKICETVDPEGKFLYLGSAGVPTCYEFDIDNKHEKE
        IIAKEVEEFIKYFNQGVENSKMNGNKKICDGEKDMNNLEILLENKLDITIGPKY
        NEKDEYLNENCKIIKYPNRIVCIGEGIDFDRLYCSKTIQKIFIFQKLVGMFN
        LPMFLHNRNSETFKVIDIYKFLFXNGGVVHSFTDKEDIVHIVQNKLYIGVNG
        CSLSLNINAVKRIPLNLLLETDAPMGCVKKTASVEYIKDYERAYTNLKKIKN
        IIKDDNTIFKERNEPYNIA"
        complement(2599..2610)
        /gene="MALIP3.01"
        /note="potential splice acceptor sequence"
        complement(2742..2747)
        /gene="MALIP3.01"
        /note="potential splice donor sequence, atg/gtataa"
        complement(2849..2861)
        /gene="MALIP3.01"
        /note="potential splice acceptor sequence"
        complement(2984..2989)
        /gene="MALIP3.01"
        /note="potential splice donor sequence, aaa/gtataa"
        5005..5496
        /gene="MALIP3.02"
        5005..5496
        /gene="MALIP3.02"
        /note="MALIP3.02, hypothetical protein, len: 163 aa,
        contains possible signal sequence"
        /codon_start=1
        /product="hypothetical protein, MALIP3.02"
        /protein_id="CAB63557.1"
        /db_xref="GI:6594245"
        /translation="MKLLNRRFVVLCPILIFLFLNSVVLGNRRNRRNINFHETENAAK
        AMRLLSGEISIKLNDGDLKIKLNDKHKDSTKWDKSYSFISNLEEKYSQTDLFR
        KKOENANTWIIEDRQEFYLNDEIENIATRVLENFNNDELIIQSPKSLIDIOS
        LNN"
        8020..10389
        /note="possible cent, region of very high [A+T] content"
        14884..20352
        /gene="MALIP3.03"
        14884..20352
        /gene="MALIP3.03"
        /note="MALIP3.03, putative ABC transporter, len: 1822 aa"
        /codon_start=1
        /product="putative ABC transporter"
        /protein_id="CAB63558.1"
        /db_xref="GI:6594246"

```

```

/translation="MTTYKENVGISNKGNNKKKSCQNISFLNLSFOWIRPLINDLIK
GDICELNICRNPDPVYASKLEENLRDIEVEDSEFYSEKNSNHSFLHHCNSNDLASE
KKYVYVYHNILMSILKTEFRIILISFYLETLVTLGCKFDIYMRILEGOKIPV
YISLDFKVFESGLVYVYMHFFHFFALLHFPHLFTINKVSLMFLYKINCUSNN
NHLGNPDAYNTYRKFSQTEIDISRDLSIGNASSSSSGIKNNKNDNNKFNVEN
DYINFIKSTKMEKDSLENRSLPNVINYIMFSDVPSVTFFSYTISLNFNVFKIF
MSFYVFIKIGSNVGIAIWLISALYSAMILFEPLSFKSKYLIYRDKRIDNNHHVL
KEFLIKMNESEAFKYINFRKMKYKIRLYLSNIGVYFISISDVIDEYVIFFI
YKDLRANKKEIKETSIIMPLYVVKILSNVAPNLVNMVMEGIVIKRLNVIYNDH
LYNDIKNFWRYRVEDYINIVDKTFLQENITSDHDTGSHLKLKLVKKNLITN
MKYFFYHKMNYHNKIINQIISGLLNKNDVDDNTKKICFOEHSKNSNTINTNSHIHE
KKEYENIHNSSTMSNEKKEKKNNYIILKLECSGLSYDNKDCNDHILKNINFN
LKRNSLAIIIGNVGSGSAFHSIQLSGLLNDSTIEHGDWKYINDDHNLKSGOKVRLCLA
NIRSMILFNGEYNPLIVKYITLQSELLNDSTIEHGDWKYINDDHNLKSGOKVRLCLA
RALYEHVIMHKLCTDYEKLIQNEILDKLINNKITSSYNNKSKLKNVYNIPNEN
YLOKLDMDNNFYLYLDDITFSDPSISKKIFSNLCKEDNISFKDNCSPFIISNKS
TLDFLEILDNVOYEYINIEIQTDLTKYRGNSEYMEKNLNLITKESHGYSNLNT
IDYTRIKLDEVELNHNKSMYKAYFYKNGTESVFEIDSINKEYIKMKKKNY
KKEHMKNNKNDNNNNKNDHINIMNDHNRNYNDINLGNSTDDSPVSSLGNE
YLDOTYSNDSKDEIVKPLYKDHHEFNKSSMPKVSNNMNNPSNFYEDNSSS
PKGSISLETYLWYFQGVGLLTSVVFMLISIEDEIKFVELTMSIISKNNKHS
TILQKQVRIEYFVPIISLVSIGCFSMIIIGNTSAIKVHNHILYSILNAPLYIF
YNNLGNIIINFIIDISAFQFLKRIYKAFFIFRCLISLLIIMRIDCIFIPFPY
ILITYFFVFRFSRGCQAQRLYSCHTPLCNIYSNALSNGKNIINIKKNTYHLDVYE
HVINERTSYFFKWLINIWASLYIKIFILLTLYIIMPHLYASGIKLYKKNVRI
LSTLGYCISFARGVILKFLCDYTHEKEMCCVORLEEFKISNKENASMKENEL
KITVNGTYKYIDEEPSLNIMYALKNQKIGVSGAGSKYDGKNTLSILGLINLSQK
KITVEGDRTYRNKGEDSIIGLAQSFYFNWNRITFDYNNFDDDELVAHLKN
GINKGLDLYKMHKQNSYKRIQTSKVINQSNONTILLTNDCIYRLSLVLYLN
RHVKYIILIDIPILFNLNKSHDELNSFLIGKAKSFNYIIRNHPNNTVLIIISHRANT
LSCDXYIVLRKGEITYRCSYEDVKTQSELSHLEMD"
23896..31533
        /gene="rRNA"
        /note="rRNA region containing small subunit, 5.8S and large
        subunit rRNA genes and spacer regions"
        23896..31533
        /gene="rRNA"
        complement(31966..32775)
        /gene="MALIP3.04"
        complement(join(31966..32476,32675..32775))
        /note="MALIP3.04, conserved hypothetical membrane protein,
        len: 203 aa, similarity: P. falciparum chromosome 2,
        PF0110W, O96126 predicted integral membrane protein (255
        aa), fasta scores: Opt: 335, E(): 4.9e-15, (36.1% identity
        in 191 aa overlap)"
        /codon_start=1
        /product="conserved hypothetical membrane protein,
        MALIP3.04"
        /protein_id="CAB63559.1"
        /db_xref="GI:6594247"
        /translation="MKKSTFINVTILLFLLLFLTYVYDTSKTFNNNKKIDIN
        RFKRIIAESEQYWEEDFCILNEELIRPHNDSPYLPHEYENIDELNSINS
        TKIWKETTKKRONYKEDTNNHNRDRFMHYKWKWANYLYKVKHLINIKLDTNPI
        HDKEETITTWIKIQEDIEYFLFNQVLEWLILILEFYKKE"
        complement(32477..32486)
        /gene="MALIP3.04"
        /note="potential splice acceptor sequence"
        complement(32669..32674)
        /gene="MALIP3.04"
        /note="potential splice donor sequence, aaa/gtataa"
        36657..37343
        /gene="MALIP3.05"
        join(36657..36743,36864..37343)
        /gene="MALIP3.05"
        /note="MALIP3.05, hypothetical protein, len: 188 aa"
        /codon_start=1
        /product="hypothetical protein, MALIP3.05"
        /protein_id="CAB63560.1"
        /db_xref="GI:6594248"
        /translation="MRIMKNSGIFIKLLICISFICVPECFNCKMISYRKOLLWYSEN
        CFNYSIDRSLAEGSSSEKTKVDIPNIELKSLNIEYEEKMKKEIVGSPMDNNLN

```







FEATURES	source	Location/Qualifiers
		1..130281
		/organism="plasmodium falciparum"
		/db_xref="taxon:5833"
		/chromosome="12"
		/clone="3D7"
BASE COUNT	52250 a	11780 c 11855 g 53996 t 400 others
ORIGIN		

Query Match	4.1%	Score 89.6;	DB 60;	Length 130281;
Best Local Similarity	47.3%	Pred. No. 0.0022;		
Matches 534;	Conservative	0;	Mismatches 574;	Indels 20; Gaps 8;









release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector phase etc.

Order of segments is not known; 800 n's separate segments.

\* NOTE: This is a 'working draft' sequence.

\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES	source	location/Qualifiers
		1. 80518
		/organism="Plasmodium falciparum"
		/strain="3D7"
		/db_xref="taxon:5833"
		/chromosome="13"
BASE COUNT	29805 a 6650 c 7591 g 27672 t 8800 others	
ORIGIN		
Query Match	4.1%	Score 88.2; DB 31; Length 80518;
Best Local Similarity	44.8%	Pred. No. 0.0038;
Matches 464; Conservative	0; Mismatches 563; Indels	8; Gaps
Qy 109	ataaaattttcacattaaattagaaattttttttat-caataaataataattattagttt 167	
Db 51967	ATAAATATTATATATATTATATTATATATATATATATATATATATATATATATTT 52026	
Qy 168	tattagaaattatttagaaaaatttgaaatcccggaatttcctctctcttcctcgctat 227	
Db 52027	TTTTTTTTTTTATTCCAATAAATTAATAATATATATACCTCATGATAATATATCATTT 52086	
Qy 228	tcatacttttctaaccaaaaccaacttatatgtctctccaattagaacctgaaattatta 287	
Db 52087	ATACAATAAAAGAAAGAAAATAATATATATATATATATATATATATATATATATATA 52146	
Qy 288	attataataaacgaaaaacaatttggatcaattc----atacacatgcttagtaataaa 343	
Db 52147	ATAAATATATAAANAATTAATCTTAATAATTTTAAATAATATATATATATATATATAT 52206	
Qy 344	aatcgataataatttgataaactcgaaaaagattttacaaatatcttcagaaaaaatt 403	
Db 52207	ATACTTATTATATATAAATAATATATAAATAATTAATCTTAATAATTTTAAATATA 52366	
Qy 404	ataacaaaatttgcgttttcaatggttgcgtgaggagatttggcacctacagaaact 463	
Db 52267	TATATATTAATATATATATATATTTTATAATACTTATAAANAATAATTAATCTTAGTTGT 52326	
Qy 464	ctctacggaccattcttgcacttcaactaaaacgatggtcagaatttgggtgggatitta 523	
Db 52327	TTTTATCGTGACCTTATGTCACATATATAGGGAGCCTACATTTTTTTCCTTTAAACTTA 52386	
Qy 524	tattcaagcatatccctttcgaactctacttacttgcgtgcgttcggtaatcggtaaac 583	
Db 52387	CAAAATTTTATACGTACACATAAATTTTATAATAATAAATAATAATAATATATATATA 52446	
Qy 584	attgacatttcaaaatcatatttaccctcaaacagtaaaatttgaaggacaaaaataata 643	
Db 52447	TATATATAGTATATATAATATATCTTTATACAATATTATGTGTAATCATTAATGAACATA 52506	
Qy 644	tttttcaaaatttgatagactatttttttttgaatttggacgaacccaaacacgatttat 703	
Db 52507	TTTATATATAATAATATTAACATATATATTATATATATATATATAATAATAATAATTTTA 52566	
Qy 704	cctgaatttttaggaaccacagatgtaactaaaccaaattttatttttttctcaaaaacaaa 763	
Db 52567	ATAATTTCAAAATAAATTAATCTCGGAAAAATATCTGACTTATTTTATAAATATATATATA 52626	
Qy 764	attcatggcagcatgctcagcccatgaaaaaaccttataaaaaataatctcacacattga 823	
Db 52627	TTATATATATATAATATATATGTTATAATATATATATAATAAATAAATTTCTCATGTAA 52686	
Qy 824	ccattgaaaagtctgtctcccatcgggttaaccagatcaaacctcacatcccaaacataacat 883	

Qy	402	ttataacaaatttgcgtcttcacaggtctgagggagatttggccactatagaa	461
Db	119379	TTTATATTTTATTTATTTATTTATTTAGTATTTCTTTCTTTTGCATTTTAAAGATTGTAA	119438
Qy	462	ctctccacggaccattctcttgcactcacaacgcatggcagaaattggcgggagattt	521
Db	119439	ACGTTTACATTTTATATTAATGTATTTATTTTATAAGGATATACTTATATATT - ATATTA	119496
Qy	522	tatatcaagcatatccctctcaaaacttctacttactctgctgogttcggtcaatcggta	581
Db	119497	TATATATATATATATATTTTATTTTATTTTATTTTGTGTAGACATACTACTATTTTA	119556
Qy	582	acattagactttcaaaactcatttttaaccctcaaacagtaaatcttggaaggacaaaataa	641
Db	119557	TTATATTTTATATAAAAATTCGTTAAATAATTTATATATATATATATATATATTT	119616
Qy	642	tattttcaaatgtatagactatttttttctgtaatttgacgaaccaaaaccgagattt	701
Db	119617	TATATATTATA-TGGAAAGATACTTCCATATATGTAAATATATATATAATAATATAAATAG	119675
Qy	702	atctgaatttttaggaacc-----acagatgtaactaaaccaaatatttatttatttcta	756
Db	119676	ATCTGTATTTTATTAATCCGTTTAAATATATATTATTTAAATAATTAATGCTATATAAA	119735
Qy	757	aaacaaatttcatggcagcatgcctccagcccatgataaaaccccttataaaaaatctcac	816
Db	119736	ATAGAGGGTTTCTCTTA-AGCAAAATTTCTATATAAAATATAGTGGTGATATAAATAAA	119794
Qy	817	acattgaccttgaaagttcgttctccactgggtaacccagatcdaactcacatccaac	876
Db	119795	ATATAAATATATATATATATTTTCTACGGTGGTATTCCAAAGCAAAAGCTTAAACAAAA	119854
Qy	877	ataacatggatctctctaccatcatcactaatatttgggtcaaatcatcaatcatta	936
Db	119855	TATATATATATATATATATATATATATATATATATATAAATATATATTTTTTAAATTA	119914
Qy	937	tttttaagatattaataagaaattaaaagatttttttaaaaaatgatacaaatcatatt	996
Db	119915	TGTCGATTGAAAAAATAATAGTAGTCTCTACTTTTAAAAATATAAAATTTATTTAAATAATA	119974
Qy	997	attcatgattttccacatttattgatttttgataataaatatttttttttcaattctctaaa	1056
Db	119975	AATATATATATATATATAATATTTTATGCAATTAATCCTTTGTCATGTTTATTTTAAATATA	120034
Qy	1057	aaatgttgcagacacttatta	1078
Db	120035	ATAAAATATATGAAAAAATTTTA	120056

RESULT 15	
PFMAL13PA	
LOCUS	PFMAL13PA 80518 bp DNA HTG 19-AUG-1999
DEFINITION	Plasmodium falciparum chromosome 13 strain 3D7, *** SEQUENCING IN
	PROGRESS ***, in unordered pieces.
ACCESSION	AL109815
VERSION	AL109815.2 GI:5763806
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	malaria parasite P. falciparum.
ORGANISM	Plasmodium falciparum
	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE	1 (bases 1 to 80518)
AUTHORS	Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M., and Barrell, B.
TITLE	Direct Submission
JOURNAL	Submitted (11-AUG-1999) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT	On Aug 24, 1999 this sequence version replaced gi:5731892. For more information about this sequence or the Malaria Project, see <a href="http://www.sanger.ac.uk/Projects/P_falciparum">http://www.sanger.ac.uk/Projects/P_falciparum</a> . IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 00:25:01 ; Search time 2111.72 Seconds  
(without alignments)  
3286.927 Million cell updates/sec

Title: US-09-464-528-15  
Perfect score: 1574  
Sequence: 1 atcgatagacatgtatt.....gttttgaagtataaagatg 1574

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*

44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: em\_est35:\*  
83: em\_est36:\*  
84: em\_est37:\*  
85: gb\_est48:\*  
86: gb\_est49:\*  
87: gb\_est50:\*  
88: gb\_est51:\*  
89: gb\_est52:\*  
90: gb\_est53:\*  
91: gb\_est54:\*  
92: gb\_est55:\*  
93: gb\_gss1:\*  
94: gb\_gss2:\*  
95: gb\_gss3:\*  
96: gb\_gss4:\*  
97: em\_gss1:\*  
98: em\_gss2:\*  
99: em\_gss3:\*  
100: em\_gss4:\*  
101: gb\_gss5:\*  
102: gb\_gss6:\*  
103: gb\_gss7:\*  
104: gb\_gss8:\*  
105: gb\_gss9:\*  
106: em\_gss5:\*  
107: em\_gss6:\*  
108: em\_gss7:\*  
109: em\_gss8:\*  
110: em\_gss9:\*  
111: em\_gss10:\*  
112: em\_gss11:\*  
113: gb\_gss10:\*  
114: gb\_gss11:\*  
115: em\_gss12:\*  
116: gb\_gss12:\*



117: gb-gss13:.\*  
 118: gb-gss14:.\*  
 119: gb-gss15:.\*  
 120: gb-gss16:.\*  
 121: gb-gss17:.\*  
 122: gb-gss18:.\*  
 123: gb-gss19:.\*  
 124: em-gss13:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	95	6.0	1101	122	CNS00EVL	AL069706 Drosophila
C 2	92	5.8	1101	122	CNS00EVL	AL069706 Drosophila
C 3	89.2	5.7	1101	122	CNS00EVL	AL069440 Drosophila
C 4	88.6	5.6	928	122	CNS00DKY	AL071865 Drosophila
C 5	87	5.5	987	123	CNS014PQ	AL104456 Drosophila
C 6	85.2	5.4	1101	122	CNS003BD	AL064091 Drosophila
C 7	84	5.3	1101	122	CNS0039G	AL063921 Drosophila
C 8	83.8	5.3	843	122	CNS00CS1	AL059666 Drosophila
C 9	83	5.3	1101	122	CNS0021J	AL061936 Drosophila
C 10	82.8	5.3	1101	123	CNS0161D	AL106896 Drosophila
C 11	81.6	5.2	1225	123	CNS0161D	AL106171 Drosophila
C 12	81.2	5.2	928	122	CNS00DKY	AL071865 Drosophila
C 13	81.2	5.2	1201	123	CNS0161D	AL106627 Drosophila
C 14	81	5.1	1043	123	CNS0145P	AL103735 Drosophila
C 15	80	5.1	1200	123	CNS016CO	AL106578 Drosophila
C 16	79.8	5.1	996	122	CNS00FUG	AL071063 Drosophila
C 17	79.2	5.0	1101	122	CNS001FB	AL060732 Drosophila
C 18	78.6	5.0	1101	122	CNS00VWL	AL059627 Drosophila
C 19	77.4	4.9	1101	122	CNS000B8	AL063632 Drosophila
C 20	77.4	4.9	1101	122	CNS00EVL	AL069440 Drosophila
C 21	76.8	4.9	1101	122	CNS00FVG	AL071206 Drosophila
C 22	76.8	4.9	1201	123	CNS0167M	AL106396 Drosophila
C 23	76	4.8	1101	122	CNS00B01	AL057419 Drosophila
C 24	75.6	4.8	734	122	CNS010MP	AL059163 Drosophila
C 25	75.6	4.8	1101	122	CNS0021J	AL061936 Drosophila
C 26	75.6	4.8	1101	122	CNS0039G	AL063921 Drosophila
C 27	75	4.8	1201	123	CNS0167M	AL106396 Drosophila
C 28	74.8	4.8	1101	122	CNS001VP	AL078809 Drosophila
C 29	74.4	4.7	987	123	CNS014PQ	AL104456 Drosophila
C 30	74.4	4.7	996	122	CNS00FUG	AL071063 Drosophila
C 31	74.4	4.7	1101	122	CNS003DQ	AL064580 Drosophila
C 32	74.4	4.7	1101	122	CNS00FVG	AL071206 Drosophila
C 33	74.2	4.7	1043	123	CNS0145P	AL103735 Drosophila
C 34	74.2	4.7	1225	123	CNS0161D	AL106171 Drosophila
C 35	73.2	4.7	994	123	CNS015XG	AL106030 Drosophila
C 36	73.2	4.7	1101	122	CNS001F2	AL078714 Drosophila
C 37	73	4.6	1101	122	CNS003BD	AL064091 Drosophila
C 38	72.8	4.6	820	120	B11728	B11728 T8D18-Sp6.1
C 39	72.6	4.6	860	123	CNS018BP	AL109135 Drosophila
C 40	72.2	4.6	1101	122	CNS00EPO	AL069493 Drosophila
C 41	72	4.6	1101	122	CNS00EPO	AL070972 Drosophila
C 42	71.8	4.6	843	122	CNS00CS1	AL059666 Drosophila
C 43	71.6	4.5	1101	122	CNS0022H	AL097139 Drosophila
C 44	71.6	4.5	1248	120	B11336	B11336 F19M10-Sp6
C 45	71.4	4.5	770	113	AQ740708	AQ740708 HS_5507_A

## ALIGNMENTS

RESULT 1  
 CNS00EVL/c 1101 bp DNA GSS 04-JUN-1999  
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:  
 DEFINITION BACR29823 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

AL069706  
 AL069706.1 GI:4949849  
 GSS  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

Melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammosier in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain Y2; cn bw sp, the same strain used for the BDGP's

P1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers

1. 1101

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_lib="RPCI-98"

/clone="BACR29823"

/note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others

ORIGIN

Query Match

Best Local Similarity 6.08; Score 95; DB 122; Length 1101;

Matches 197; Conservative 100; Mismatches 270; Indels 0; Gaps 0;

QY 573 taatcggaacattagacattcaaaatcatttttaacccttaaacagtaattgaagga 632

DB 1059 TWTWTAT 1000

QY 633 caaaaataattttcaaaattgatagactattttttttgttaattgcgaaccaa 692

DB 999 WTAT 940

QY 693 accgattatccctgaatttttaggaacacacagatgaactaaacattatttttt 752

DB 939 AWWTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 880

QY 753 tctaaacaaatctcggcagcagcctcagcccatgaaacacccctataaaatcat 812

DB 879 TTTTWTAT 820

QY 813 ctacacattgaccattgaaaggttcgtctcccatgggtaaccagatcaaacatcc 872

DB 819 WATWTAT 760

QY 873 aaacataacatggatattccctaccataactaactatttttgggttaaatattatc 932

DB 759 WAT 700

QY 933 attatttttaagattcaataagaataataaaatatttttaaaaaatgataaaatta 992

DB 699 TTTTWTAT 640





BP 191 91006 EVRY cedex - FRANCE (E-mail : [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoedawa and  
Aaron Mammossier in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain Y2: cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES . Location/Qualifiers  
1. 1101  
source

BASE COUNT		ORIGIN	
395 a	120 c	103 g	334 t
		149 others	

	Query Match	5.4%	Score 85.2	DB 122	Length 1101
	Best Local Similarity	40.5%	Pred. No. 3.4e-03		
	Matches 223	Conservative	56	Mismatches 285	Indels 6
	Gaps 1				
QY	581	aacattgactctcaaaatcatcttttaaccctcaaacagtgaaatttgagcgagcaaaaata	640		
		: :     :     :			
Db	483	MAAAAAAAAAATWTTAAAAAANWAAAAAACCTTTAATAAANWAAAAAANWAAAAAANA	542		
		: :     :     :			
QY	641	atattttcaaatgttagactatctttttttgtatattgacgaacccaaccagatt	700		
		: :         :			
Db	543	AWTTTTTTTWWAMTTTTATAMATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	602		
		: :         :			
QY	701	tatctgaaattttaggaaaccacagatgtcaactaaccaatcttattctttcttaaac	760		
		: :			
Db	603	MATTTTTTTTTTTTT-----TTTTTTTTTTTTTTTTTATTTTTTWTATNTTWTATTTTAA	656		
		: :			
QY	761	aaaatttcaggcagcatgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	820		
		: :			

[illegible][illegible]

**D**b      897    AATTAAATTAATAAAAAAATTAAATATAAAAATTAATTAATTTTNTYTAWAAAAAYWAT    956

Qy 1061 gtgtgaagacattattagacatagctctgttctgtttacaaaagcattcatcattta 1120

Dd 957 TAAWAWAATATTTTATATATATATATATATAATWAAAAAAAATAAATAATWAAATTNWA 1016

Qy 1121 acattataaaa 1130  
| : : ||| :

.Db 1017 AWWAYAAWA 1026

QY	606	taaccscctaacaacagtaaatttgaaggacaaaaataaatatcttttcataaaattgatagaactat	665
DB	762	HAMMMHNAHYAALAAAWAATTHNHNHTTHNMHTTUMYUUYUTCCYMCSTYHCWNUYT	821
QY	666	tcttttttgtaattgacgaacaaaacsacagatttatctcgtaaatttttaggaacacacaga	725
DB	822	AUTCSTWTNNMNTWTGHYNHTTNWNHTTTGNAWHNHTTCWMMNHATTWTWATHCWAOMT	881
QY	726	tgtaacctaacaacaaattattattttctttaaacaacaaattcatgcggcacgactcgacg	785
DB	882	MNNHNNMNNHHNNHNSACHNNTTSMCSNNHMCTCHNHNHTYHMTSCHMWMNNHNNHNNHW	941
QY	786	cscatgaacaaaacctataaaaaatatcatcacatgacattgaaagtcttgtctctccc	845
DB	942	MATWMTTWTMMIMSSMNHNCHNVHMNMHMYUCSCYUUCTCHTATTHYHNMCTCYNYCT	1001
QY	846	atogggtaacacagatcaaacatcacatccaaacataacatgagatattctcttccaatcata	905
DB	1002	WHTYTAYWWATTAHAHTTATMMNNHNNHNAHTWWWMMWATAWNACTCHNHTWTYHTH	1061
QY	906	ctaattatttgggttaaatataatcatcattttt	941
DB	1062	CTWYUYNHTTWMAWMAWNNHNNHNTAHYHNAHNCWY	1097

RESULT 8  
 CNS00CSI/ 843 bp DNA GSS 04-JUN-1999  
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC  
 FLYR26H19 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
 ACCESSION AL059666  
 VERSION AL059666  
 KEYWORDS GI:4947129  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 843)  
 REFERENCE Genoscope.  
 AUTHORS Direct Submission  
 TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequenc  
 JOURNAL Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 COMMENT - Web : www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACS. For further information  
 please see http://www.fruitfly.org The BDGP Drosophila  
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
 Aaron Hammoser in Pieter de Jong's laboratory in the Department  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPCI-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from partial  
 isogenic strain y2; cn bw sp, the same strain used for the BDGP  
 pl and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be  
 found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

FEATURES	Location/Qualifiers
source	1..843
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone_lib="RPCI-98"
	/clone="BACR26H19"
	/note="end : TET3"
BASE COUNT	131 a 102 c 259 g 141 t 210 others
ORIGIN	

Query Match 5.3%; Score 83.8; DB 122; Length 843;

```

Best Local Similarity 39.2%; Pred. NO. 6e-05;
Matches 122; Conservative 64; Mismatches 125; Indels 0; Gaps 0;

Qy 851 taaccagatcaaacacacatacaatacaatgagatctctcttcaacatacactaat 910
  || || || || || || || || || || || || || || || || || || || ||
Db 833 TATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAW 774
  || || || || || || || || || || || || || || || || || || || ||
Qy 911 ttttttgggttaataatcattattttaaagataataataaagaaatgattt 970
  || || || || || || || || || || || || || || || || || || || ||
Db 773 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 714
  || || || || || || || || || || || || || || || || || || || ||
Qy 971 ttaaaaaaatgataaaatattattattcattgatttttccatacaattgattttgataat 1030
  || || || || || || || || || || || || || || || || || || || ||
Db 713 ATATTATATATATATATATATATATATATATATATATATATATATATATATAT 654
  || || || || || || || || || || || || || || || || || || || ||
Qy 1031 aaatatattttttttaaattcttcaaaaaatgttgcagaacactttattagacatagctctg 1090
  || || || || || || || || || || || || || || || || || || || ||
Db 653 AWTWTWTWTATWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATA 594
  || || || || || || || || || || || || || || || || || || || ||
Qy 1091 tctgtttcaaaaagcattcattcattcaatacaataaaatatttaatacactaacagtag 1150
  || || || || || || || || || || || || || || || || || || || ||
Db 593 AAATWATATATAAAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATA 534
  || || || || || || || || || || || || || || || || || || || ||
Qy 1151 aatcttcttgt 1161
  || || || || || || || || || || || || || || || || || || || ||
Db 533 WAATWATWAT 523

RESULT 9
CNS0021J/C CNS0021J 1101 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL061936
VERSION AL061936.1 GI:4940214
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammos in Pletier de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone-lib="RPCI-98"
/clone="BACR05N11"
/note="end : TET3"
/feature="end : TET3"

BASE COUNT 631 a 7 c 28 g 289 t 146 others
ORIGIN

```

```

Query Match 5.3%; Score 83; DB 122; Length 1101;
Best Local Similarity 40.2%; Pred. No. 7.9e-05;
Matches 210; Conservative 61; Mismatches 245; Indels 6; Gaps 1;

Qy 740 tatttatttttcttaaaacaaattcttcagcagcatgcctcagcccatgaaaaaac 799
  || || || || || || || || || || || || || || || || || || || ||
Db 1043 TTTTTTTTTTTTTTTTNNMAAATATTHMCACATTTTTCATNCYTCWTATTTTTCMTWH. 984
  || || || || || || || || || || || || || || || || || || || ||
Qy 800 cttataaaatactcacacattgacattgaaaagtctctccatgggtaaacagat 859
  || || || || || || || || || || || || || || || || || || || ||
Db 983 TTTTAMMATTMAAMMATTATYMTCTTACHATTYTTAAACMYCMMYTAMCMHMCATTC 924
  || || || || || || || || || || || || || || || || || || || ||
Qy 860 caactccatcccaacatacattgatatctccctaccaatcataactattcatt---- 914
  || || || || || || || || || || || || || || || || || || || ||
Db 923 MCAWTTTTTAACTTTAAACACATATTTAAATSAATTTTAAATTTTAAATAAATA 864
  || || || || || || || || || || || || || || || || || || || ||
Qy 915 -ttgggttaaatcattcatttttaagataataatttaagaaattaaaagattttt 973
  || || || || || || || || || || || || || || || || || || || ||
Db 863 ATTTAAAAAAAWTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 804
  || || || || || || || || || || || || || || || || || || || ||
Qy 974 aaaaaaatgataaaatattattattcattgatttttccatacattgattttgataataa 1033
  || || || || || || || || || || || || || || || || || || || ||
Db 803 ATTAATAAAATTAATAATTAATAATTTTAAATTTTAAATTTTAAATTTTAAATTT 744
  || || || || || || || || || || || || || || || || || || || ||
Qy 1034 tatattttttttaaattcttaaaaaatgttgcagaacactttattagacatagctgttc 1093
  || || || || || || || || || || || || || || || || || || || ||
Db 743 TWTTTTAAATTAATAATTTTAAAWAAAAAAATTAATAATTTTAAATTTTAAATTT 684
  || || || || || || || || || || || || || || || || || || || ||
Qy 1094 tgtttacaagaagcattcattcattcattcattcattcattcattcattcattcattcatt 1153
  || || || || || || || || || || || || || || || || || || || ||
Db 683 WATWATAAAWAAWTTWTATWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 624
  || || || || || || || || || || || || || || || || || || || ||
Qy 1154 ctctctggagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1213
  || || || || || || || || || || || || || || || || || || || ||
Db 623 TTTTWTWTWATTAATAATTAATAATTTTAAATTTTAAATTTTAAATTTTAAATTT 564
  || || || || || || || || || || || || || || || || || || || ||
Qy 1214 aaccagaagacaaataaagaatgtagcaacacactttattagacatagctgttc 1255
  || || || || || || || || || || || || || || || || || || || ||
Db 563 TTAWAAAAAAWAAAAAAWAAAAAAWAAAAAAWAAAAAAWAAAAAAWAAAA 522

RESULT 10
CNS016LI/C CNS016LI 1101 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106896
VERSION AL106896.1 GI:5624374
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CSHP (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelBAC11.
Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"

```



## FEATURES





Result No.	Score	Query Match	Length	DB	ID	Description
1	57.2	3.6	3095	7	5231168-1	Patent No. 5231168
2	56	3.6	19124	4	US-08-487-826B-13	Sequence 13, Appl
3	54.2	3.4	11864	4	US-08-731-722-5	Sequence 5, Appl
4	53.8	3.4	6243	4	US-09-056-075-1	Sequence 1, Appl
5	53.2	3.4	3138	1	US-07-867-106-4	Sequence 4, Appl
6	52.6	3.3	665	4	US-08-883-795A-36	Sequence 36, Appl
7	52	3.3	8920	3	US-08-446-835A-1	Sequence 1, Appl
8	51.4	3.3	19124	4	US-08-487-826B-13	Sequence 13, Appl
9	51.2	3.3	665	4	US-08-883-795A-36	Sequence 36, Appl
10	51.2	3.3	3933	4	US-08-731-722-3	Sequence 3, Appl
11	51.2	3.3	3933	4	US-08-731-722-3	Sequence 3, Appl
12	50.6	3.2	8920	3	US-08-446-835A-1	Sequence 1, Appl
13	50.2	3.2	473	2	US-08-764-100-16	Sequence 16, Appl
14	50.2	3.2	4970	2	US-08-764-100-14	Sequence 14, Appl
15	50.2	3.2	4970	2	US-08-764-100-20	Sequence 20, Appl
16	49.8	3.2	4467	2	US-08-563-907A-1	Sequence 1, Appl
17	49.8	3.2	4467	4	US-08-910-551B-1	Sequence 1, Appl
18	49.8	3.2	4467	4	US-08-909-435A-1	Sequence 1, Appl
19	49.4	3.1	1611	7	5213972-6	Patent No. 5213972
20	49.2	3.1	319	1	US-07-593-657-14	Sequence 14, Appl
21	49	3.1	4098	4	US-08-503-106-4	Sequence 4, Appl
22	49	3.1	6768	2	US-08-107-755A-1	Sequence 1, Appl
23	49	3.1	8457	1	US-07-991-867B-1	Sequence 1, Appl
24	49	3.1	8457	4	US-08-544-332-1	Sequence 1, Appl
25	48.4	3.1	1415	1	US-08-413-118-126	Sequence 126, App
26	48.4	3.1	1415	5	US-08-473-446-136	Sequence 126, App

Query Match	3.6%	Score 56;	DB 4;	Length 19124;	
Best Local Similarity	47.7%;	Pred. No. 0.012;			
Matches 164;	Conservative	0;	Mismatches 180;	Indels 0;	Gaps 0;
Qy	909	atctatttgggttaaatatcaatcattatttttaagatatattaataagaataaaaaagat	968		
Db	1264	ATTATAATATGTAAATTTATTAAATAAATATATTGTTATACATCAAGAGCTAAAGAAAAC	1323		
Qy	969	tttttaaaaaatgtacaaaattatatattcatgatttttcacattgatttttgata	1028		
Db	1324	TATACATCTGGTATCTAATAGTATATATATATAATATCTTTTTTATTAACTTCTCT	1393		
Qy	1029	ataatatattttttttaaattttcttaaaaaatgtgcaagacacttattagacatagct	1088		
Db	1384	CTTTTTTTTTTTTTAAATAATAATAAATTAATAATATATTTTTTTTCATAATTATATGAT	1443		
Qy	1089	tgttctgtttcacaaaagcattcatcattttaaatacatataaaaaattttaatactaacagt	1148		
Db	1444	TTAGTATTTTAAATAATAATAAATCTTTTTTAAAAAAGCTTCAAAAGCATTTTTCATATAAATA	1503		
Qy	1149	agaactctcttctgtagtcggtgtgggagtaggcaacctggcattgcaacgagagaaagaga	1208		
Db	1504	ATATTAATATTAGTAACCACTAGATAAATTAGAGAGAAAGGTAGACATACCAAAAAA	1563		
Qy	1209	gtcagaaccgagagacaaaataaaaaagtatgcacacaaacaatca	1252		
Db	1564	ATTAGAACAAAAGAAATATTACAAAAAATAATAAAATTTAAATTA	1607		

RESULT 3  
 US-08-731-722-5/c  
 Sequence 5, Application US/08731722  
 Patent No. 5961971  
 GENERAL INFORMATION:  
 APPLICANT: Martin, Frank N.  
 TITLE OF INVENTION: Biocontrol of Fungal Soilborne Pathogens  
 TITLE OF INVENTION: by Pythium oligandrum  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Saliwanchik & Saliwanchik  
 STREET: 2421 N.W. 41st Street, Suite A-1  
 CITY: Gainesville  
 STATE: FL  
 COUNTRY: US  
 ZIP: 32606-6659  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/731.722  
 FILING DATE:  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Whitlock, Ted W.  
 REGISTRATION NUMBER: 36,965  
 REFERENCE/DOCKET NUMBER: UF-161  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 352-375-8100  
 TELEFAX: 352-372-5800  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1186 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 ORIGINAL SOURCE:  
 INDIVIDUAL ISOLATE: 23-5  
 US-08-731-722-5

```

Query Match      3.4%; Score 54.2; DB 4; Length 1186;
Best Local Similarity 59.3%; Pred. No. 0.018;
Matches 112; Conservative 0; Mismatches 73; Indels 4; Gaps

```

QY	904	tactaattattttgggttaaatacttaataattttttaaagataatttaagaagaattaa 963
Db	595	TATTGAAGAATTTCCTTATAAATCTACTATTAAAGATTTTAGGATCTACTGTATAAATCCAG 536
QY	964	asgatttttttaaaaaatgtataaaaattatattcttcagatttttttcatacatatttgattt 1023
Db	535	ATCATTATTATAAATAATAAATAAAGTTT-----TTCAATATTTTATAATAAATATATTC 480
QY	1024	tgataataatatt 1083
Db	479	TTATTAGAAGTATTTTCATTTTAATTTTTTTTTTTTAAAGCTTATATATCTTTAAAAAAGATAT 420
QY	1084	agctctgtt 1092
Db	419	AAATTAATT 411

```

RESULT      4
US-09-056-075-1
; Sequence 1, Application US/09056075
; Patent No. 595368
; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric A.
; APPLICANT: Bradshaw, Marite
; APPLICANT: Rood, Julian

```

## TITLE OF INVENTION: Expression System for Clostridium

TITLE OF INVENTION: Species

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles &amp; Brady

STREET: 1 South Plackney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/056,075

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27396

REFERENCE/DOCKET NUMBER: 960296.95238

TELEPHONE: 608-251-5000

TELEFAX: 608-251-9166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 6243 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 3770..4013

OTHER INFORMATION: /note= "RP4 origin of DNA transfer (orit) from

plasmid RP4"

US-09-056-075-1

Query Match 3.4%; Score 53.8; DB 4; Length 6243;  
Best Local Similarity 50.8%; Pred. No. 0.026;  
Matches 153; Conservative 0; Mismatches 147; Indels 1; Gaps 1;

QY 125 aaattagaatttttttttatacaataaataattattatttagttttatttagaa-ataataat 183

DB 2802 AATTGAAACTTAATCTATTTTATGTTTAAATTTATTTTATTTTATTTTATTTTAA 2861

QY 184 tagaaaatttgaatcccgatttccctcttcttcttcttcttcttcttcttcttctt 243

DB 2862 CTAAATTAATGATCTTTTAAATTTTACTATTTCATTCATTCATTCATTCATTAAT 2921

QY 244 aaaccaattttatttcttcaaatagaacttgaatttgaatttgaatttgaatttga 303

DB 2922 TATTTACAAATAATTTCTTCATTTGTAATTTAGATGATTTACTAATTTTATTTT 2981

QY 304 aaacaatttggatcaattcatatcatatccttagtataaataatgcgaatttgaata 363

DB 2982 ATATATTAATAATTAATGATTAATTTATATAAATAAATAAAGGAGCTTATAAATTA 3041

QY 364 aatcgcgaataattttcaataatcttcttcagaaaaataataataataataataata 423

DB 3042 TTATTTCCAAAGATACTAAAGATTTAAATTTTCAATTTTAAACAATACTTTTGT 3101

QY 424 t 424

DB 3102 T 3102

RESULT 5

US-07-867-106-4

; Sequence 4, Application US/07867106

Patent No. 5389526

GENERAL INFORMATION:

APPLICANT: Slade, Martin B

APPLICANT: Chang, Andy C M

APPLICANT: Williams, Keith L

TITLE OF INVENTION: Improved Plasmid Vectors for Cellular

TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &amp; No. 5389526ris

STREET: One Liberty Place 46th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/867,106

FILING DATE: 19920625

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PJ 7187

APPLICATION NUMBER: PCT/AU90/00530

FILING DATE: 02-NOV-1989

ATTORNEY/AGENT INFORMATION:

NAME: Feeney, Joanne Longo

REGISTRATION NUMBER: 35,134

REFERENCE/DOCKET NUMBER: RICE-0002

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 3138 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO

US-07-867-106-4

Query Match 3.4%; Score 53.2; DB 1; Length 3138;  
Best Local Similarity 46.7%; Pred. No. 0.031;  
Matches 169; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 723 agatgttaactaaaccaattttatttttttttttttttttttttttttttttttttttttt 782

DB 2593 AGAAATAAATTTTTTTTTTTTTTTTTTTTGTCTATGACACTTTTTTTTTTTTGTCTATGACA 2652

QY 783 cagcccatgaaaaaaacaccttataaaaaatctacacattgaccattgaaaaagttcttct 842

DB 2653 GAATTCAGAAAAACACAGAAAGTTATATATTACCCCTTTTAAATTTTTTTTAAACCTTTTGA 2712

QY 843 cccatgggttaaccagatcaaacatcacatccaaacataacatgatatctctctaccacatc 902

DB 2713 AACTTTAGTAATAAGATCGATCTATCTACTTCAGTACGAACATAAATATGTATAAACCAAAA 2772

QY 903 atactaatattttgggttaataataataataataataataataataataataataataata 962

DB 2773 AAATTCATTAAAGTAAGTTATATGTTGTTATTTAAATAAATAGTTAGTTTAAATTTT 2832

QY 963 aaagatttttttaaaaaatgtataaaattatatttttttttttttttttttttttttttttt 1022

DB 2833 ATATCATTTTTTAAATAAATGAAATGTTTGAATAAATAAATTTTTTTTTTTTTTTTCAAC 2892

QY 1023 ttgataataataatatt 1082

DB 2893 GGGACCATGCTAATATCATATATGATTCATAAATTAAGTTATTAAACAAATATGTATAAAT 2952





```
Db 199 TTTAATTATAAATTTTAAATTAATAAATTTAATATAAATTTAATATAAATA 140
QY 1122 cattaataatttataactaacagtagaat 1153
Db 139 TTTAATTATAAATTTTAAATTAATAAATTT 108

RESULT 10
US-08-731-722-3
; Sequence 3, Application US/08731722
; Patent No. 5961971
; GENERAL INFORMATION:
; APPLICANT: Martin, Frank N.
; TITLE OF INVENTION: Biocontrol of Fungal Soilborne Pathogens
; TITLE OF INVENTION: by Pythium oligandrum
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/731,722
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3933 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 1986-41
;
US-08-731-722-3
Query Match 3.3%; Score 51.2; DB 4; Length 3933;
Best Local Similarity 54.3%; Pred. No. 0.076;
Matches 125; Conservative 0; Mismatches 103; Indels 2; Gaps 1;

QY 905 actaatttttgggttaataatattttaaagatatattttaaagataatttaagaaataaa 964
Db 505 ATTATTATTCTTATTAAATATTATGATTATTATTAAAGTAGAATAAGG--ATTACA 562
QY 965 agattttttaaaaaatgataaaattattattcatgatttttcatacatattgatttt 1024
Db 563 ATTTTCTGTGTTAAATAAATAAACCTTCTCTATTGAGAAATTTTATAGCTACTGTTA 622
QY 1025 gataataatattttttttaaataatattattcatgatttttcatacatattgatttt 1084
Db 623 ATTAATAAAGCTTTATATTATTTAAATCTTTTATAGAACATTTTAAATAAATATTATTAAATG 682
QY 1085 gtctgttctgttttaaaaaagcattcatcatttaataacataaaatat 1134
Db 683 GTAGAATTGAGAGAAATAAATACTTCTTAATTACTTTATATTATAATAT 732
```

```
RESULT 11
US-08-731-722-3/c
; Sequence 3, Application US/08731722
; Patent No. 5961971
; GENERAL INFORMATION:
; APPLICANT: Martin, Frank N.
; TITLE OF INVENTION: Biocontrol of Fungal Soilborne Pathogens
; TITLE OF INVENTION: by Pythium oligandrum
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/731,722
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3933 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 1986-41
;
US-08-731-722-3
Query Match 3.3%; Score 51.2; DB 4; Length 3933;
Best Local Similarity 54.3%; Pred. No. 0.076;
Matches 125; Conservative 0; Mismatches 103; Indels 2; Gaps 1;

QY 905 actaatttttgggttaataatattttaaagatatattttaaagataatttaagaaataaa 964
Db 3429 ATTATTATTCTTATTAAATATTATGATTATTATTAAAGTAGAATAAGG--ATTACA 3372
QY 965 agattttttaaaaaatgataaaattattattcatgatttttcatacatattgatttt 1024
Db 3371 ATTTTCTGTGTTAAATAAATAAACCTTCTCTATTGAGAAATTTTATAGCTACTGTTA 3312
QY 1025 gataataatattttttttaaataatattattcatgatttttcatacatattgatttt 1084
Db 3311 ATTAATAAAGCTTTATATTATTTAAATCTTTTATAGAACATTTTAAATAAATATTATTAAATG 3252
QY 1085 gtctgttctgttttaaaaaagcattcatcatttaataacataaaatat 1134
Db 3251 GTAGAATTGAGAGAAATAAATACTTCTTAATTACTTTATATTATAATAT 3202

RESULT 12
US-08-446-855A-1
; Sequence 1, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
```

APPLICANT: Peters, Dirk  
 APPLICANT: Goldbach, Robert W.  
 TITLE OF INVENTION: Improvements in or Relating to Organic  
 TITLE OF INVENTION: Compounds  
 NUMBER OF SEQUENCES: 27  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sandoz Agro, Inc.  
 STREET: 975 California Avenue  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94304  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/764,100  
 FILING DATE: 06-DEC-1996  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/214,064  
 FILING DATE:  
 APPLICATION NUMBER: US 08/032,235  
 FILING DATE: 17-MAR-1993  
 APPLICATION NUMBER: GB 9206016.9  
 FILING DATE: 19-MAR-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 577300ris, Allen E.  
 REGISTRATION NUMBER: 34,490  
 REFERENCE/DOCKET NUMBER: 137-1061  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 354-3592  
 TELEFAX: (415) 857-1125  
 INFORMATION FOR SEQ ID NO: 16:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 473 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

```

; INFORMATION FOR SEQ 10:
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 473 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;
US-08-764-100-16

Query Match          3.2%; Score 50.2; DB 2; Length 473;
Best Local Similarity 52.1%; Pred. NO. 0.09;
Matches 112; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY  922 aaatattaatcattttttaagatattaatgaagaattaaagaatttttttaaaaaaat 981
    ||| || || || || || || || || || || || || || || || || || || ||
Db  236 AAAACCCAAAAAATTTTGTGTAATAAATAAGGCTCCGCCAGATTGTGCTTAAGACC 295

QY  982 grataaaattatatttcattgttttttcacatttgatttgataataatatatttt 1041
    || || || || || || || || || || || || || || || || || || || ||
Db  296 TTTTATTATTGTTTTTATACATTATTATTGTTTGTGATTATTTTATTATTATTATTA 355

QY  1042 ttttaatttttttaaaaaattgtcgagacacctattagacatagctgtgtctgtttaca 1101
    || || || || || || || || || || || || || || || || || || || ||
Db  356 TATTTTTATATAGTTGCTTAATTACACTTATTTAGACAAATAATTTATTTTGATTA 415

QY  1102 aagcattcattcatttaatacatataaaaaatattt 1136
    || || || || || || || || || || || || || || || || || || || ||
Db  416 CAATCATCTGCGCTATTTTTAATTAAAAACACATTT 450

```

RESULT 14  
US-08-764-100-14  
; Sequence 14, Application US/08764100  
; Patent No. 5773700  
; GENERAL INFORMATION:  
; APPLICANT: van Grinsven J., Martinus Q.  
; APPLICANT: De Haan, Petrus T.  
; APPLICANT: Gielen L., Johannes J.



APPLICANT: Peters, Dirk  
 APPLICANT: Goldbach, Robert W.  
 TITLE OF INVENTION: Improvements in or Relating to Organic  
 TITLE OF INVENTION: Compounds  
 NUMBER OF SEQUENCES: 27  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sandoz Agro, Inc  
 STREET: 975 California Avenue  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94304  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: IBM PC compatible  
 SOFTWARE: Patent in Release #1.0, version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/764,100  
 FILING DATE: 06-DEC-1996  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/214,064  
 FILING DATE:  
 APPLICATION NUMBER: US 08/032,235  
 FILING DATE: 17-MAR-1993  
 APPLICATION NUMBER: GB 9206016.9  
 FILING DATE: 19-MAR-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 5773700rls, Allen E.  
 REGISTRATION NUMBER: 34,490  
 REFERENCE/DOCKET NUMBER: 137-1061  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 354-3592  
 TELEFAX: (415) 857-1125  
 INFORMATION FOR SEQ ID NO: 20:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4970 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-764-100-20

Query Match 3.2%; Score 50.2; DB 2; Length 4970;  
Best Local Similarity 52.1%; Pred. No. 0.12;  
Matches 112; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

922	aaataattaatcattattttttaagatataataataagaataattaaagaagatttttttaaaaaaat	981
QqY		
937	aaata	986
QqY		
3738	AAACCAAAAAATTTTTTTTGTAAATAAATAAGCGTCGCCAGATTTTGGCTAAGACC	3679
Ddb		
982	gtataaaaattattattcatgattttttcatacatattgatttttgataataaatatatttt	1041
QqY		
3678	TTTTTATTGTGTTTTTATACATTTATTGTGTTTGTGATTTTTTATTATTATTATTTTA	3619
Ddb		
1042	ttttaattcttcaaaaatgftgcaagacactatttagacatagcttgcttcgtttaca	1101
QqY		
3618	TATTTTTTATAGTTTGCTTTATTACACTTATTTAGACAAATTAATTTATTTCATTA	3559
Ddb		
1102	aaagcattcatcatttaatacatataaaaatattt	1136
QqY		
3558	CAATCATCTCGCTTATTTAATTATTAAACACATTT	3524
Ddb		

Search completed: September 2, 2000, 02:00:43  
Job time: 8407 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 02:02:54 ; Search time 161.21 Seconds  
(without alignments)  
2442.791 Million cell updates/sec

Title: US-09-464-528-15  
Perfect score: 1574  
Sequence: 1 atcgatagacatgtatt.....gttttgaagtataaagatg 1574

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result NO.	Score	Match	Length	ID	Description
1	63	4.0	3101	1 Q02047	Sequence encoding
2	60.2	3.8	9789	1 T41852	cDNA encoding Plas
3	59.6	3.8	1864	1 Q78892	Aspergillus nidula
c 4	59	3.7	1864	1 N71405	Sequence of ANS-1
5	58.8	3.7	53585	1 X20251	Borrelia burgdorfe
6	58	3.7	110000	1 X20248_05	Continuation (6 of
c 7	58	3.7	116277	1 X20249	Borrelia burgdorfe
8	57.2	3.6	3095	1 Q03875	Sequence encoding
9	57	3.6	6124	1 Q03568	Sequence encoding
c 10	56.4	3.6	5849	1 V33135	Plasmodium berghei
c 11	56.2	3.6	2418	1 Q27886	P.falciparum GBPI3
12	56	3.6	1864	1 N71405	Sequence of ANS-1
13	56	3.6	4590	1 N60472	Sequence encoding
14	56	3.6	19124	1 T72882	Plasmodium var-7 g
c 15	55.8	3.5	2104	1 Q25273	Sequence encoding
16	55.6	3.5	1671	1 Q24134	50 kD subunit of S
c 17	54.4	3.5	4590	1 N60472	Sequence encoding
c 18	54.2	3.4	1186	1 V29580	Pythium oligandrum
c 19	53.8	3.4	605	1 T31530	Human 3' apolipop
20	53.8	3.4	5760	1 N50530	Sequence encoding
21	53.6	3.4	2503	1 Q53480	pNPX30 xylanase cD
c 22	53.6	3.4	3775	1 V74549	Staphylococcus aur
c 23	53.6	3.4	6124	1 Q03568	Sequence encoding
c 24	53.4	3.4	1393	1 N60225	Sequence of autono
c 25	53.2	3.4	53585	1 X20251	Borrelia burgdorfe
c 26	53.2	3.4	110000	1 X20248_03	Continuation (4 of
27	52.6	3.3	605	1 T31530	Human 3' apolipop
c 28	52.6	3.3	1470	1 O55185	MS-Lel610 Vector
29	52.4	3.3	26811	1 X20253	Borrelia burgdorfe
30	52.4	3.3	110000	1 T58840_3	Continuation (4 of
31	52.2	3.3	1826	1 V37413	Orpinomyces cellul
32	52.2	3.3	1826	1 V29477	Orpinomyces cellul
c 33	52	3.3	8920	1 Q62924	Carbamoyl-phosphat

34 51.8 3.3 3975 1 N81157 Malaria-specific g  
35 51.8 3.3 3975 1 Q22999 SERP gene, Recombi  
c 36 51.8 3.3 26811 1 X20253 Borrelia burgdorfe  
37 51.8 3.3 110000 1 V21209\_06 Continuation (7 of  
38 51.6 3.3 3138 1 Q11712 Shuttle vector PMU  
c 39 51.4 3.3 9789 1 T41852 cDNA encoding Plas  
c 40 51.4 3.3 19124 1 T72882 Plasmodium var-7 g  
c 41 51.2 3.3 783 1 X20361 Borrelia burgdorfe  
42 51.2 3.3 1711 1 V33136 Plasmodium berghei  
43 51.2 3.3 3933 1 V29578 Pythium oligandrum  
c 44 51.2 3.3 3933 1 V29578 Pythium oligandrum  
45 51 3.2 110000 1 T58840\_4 Continuation (5 of

## ALIGNMENTS

RESULT 1

Q02047

ID Q02047 standard; DNA: 3101 BP.  
AC Q02047:  
DT 10-MAR-1993 (revised)  
DE Sequence encoding a circumsporozoite (CS)-related protein of  
DE Plasmodium falciparum.  
KW Plasmodium falciparum; malaria; vaccine; ss.  
OS Plasmodium falciparum T.9/96.  
FH Key Location/Qualifiers  
FT 312..1988  
FT /\*tag= b  
FT /product=CS-related protein.

WO9001496-A.

22-FEB-1990.

04-AUG-1989; G00895.

12-AUG-1986; GB-019209.

PA (REEX-) Res Exploitation Ltd.

PI Robson K, Hall J, Hallows J, Sadler R;

DR WPI: 90-083478/11.

P-PDSB: R05427.

PT DNA cloned from plasmodium falciparum - used to prepare polypeptide(s)  
used in vaccines for persons suffering from or at risk from malaria.  
PS Disclosure; Formula 1; 49pp; English.  
CC The CS protein is produced during the merozoite stage  
of the malaria parasite. The CS-related protein can be used to produce  
vaccines. The Abs can be used for the purificn. of the polypeptides.  
CC The vector may be a recombinant virus, esp. baculovirus or the peptide  
can be chemically prepared.

CC Fragments of the peptide preferably contain a conserved sequence,  
e.g. residues 244 to 291 and more particularly a polypeptide selected from  
the following group:

CC A) WDEWSPCSVTGKGRSRKR

CC B) WDEWSPCSVTGKGRSRKR

CC C) WDEWSPCSVTGKGRSRKR

CC D) PCSVTGKGR

CC E) WSPCSVTGCG

CC See also Q02049.

SQ Sequence 3101 BP; 1293 A; 350 C; 407 G; 1051 T;

Query Match

Best Local Similarity 4.0%; Score 63; DB 1; Length 3101;

Matches 427; Conservative 0; Mismatches 510; Indels 18; Gaps 4;

QY 108 atataaaattttacatttaataagaattttttttttatcaataataattattattatttt 167

DB 2095 AATAAAGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2154

QY 168 tattagaataatttaataagaatttttttttttttttttttttttttttttttttttt 227

DB 2155 AATACAAATATATATTTGTATATAAATAAATAAATAAATAAATAAATAAATAA 2214

QY 228 tcatcattttcaccac 287



[illegible]

RESULT 4  
N71405/c  
ID N71405 standard; DNA; 1864 BP.

OS Mucor miehei.  
PN EP-215594-A.  
PD 25-MAR-1987.  
PF 27-AUG-1986; 306624.  
PR 29-AUG-1985; US-771374.  
PR 07-JUL-1986; US-882224.  
PA (GENE-) GENENCOR INC.  
PI Cullen D, Gray GL, Hayenga KJ, Lawlis VB;  
PD WPI; 87-095049/14.  
PT New DNA sequences for expressing polypeptide in filamentous fungi  
PT - with secretion of prod. from the cells, and new vectors and  
PT transformed fungi.  
PS Example; Fig 13; 45pp; English.  
SC A DNA sequence coding for a heterologous polypeptide which can be  
CC expressed in and secreted from filamentous fungi is claimed. Pref.  
CC the DNA sequence codes for bovine preprochymosin, M. melnei  
CC preprocarboxyl protease or A. niger preproglucanase. Also new  
CC are vectors consisting of the DNA sequence plus an operably-linked  
CC signal sequence. The vectors may also include a sequence which  
CC increases transformation efficiency, e.g. ANS-1.  
SQ Sequence 1864 BP; 786 A; 210 C; 44 G; 732 T;

```

Query Match.      3.78; Score 59; DB 1; Length 1864;
Best Local Similarity 42.8%; Pred. No. 0.09;
Matches 452; Conservative 0; Mismatches 583; Indels 22; Gaps 5;

Qy 126 aattcagaattttttatccaataataattattattagttattattagaataattaata 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1554 ATTTCATAATNTNTTAGGAAGNTATCTATTCTCTTAAATCAATCTTAATAATTTTTTATA 1495

Qy 186 gaaaattttgaatcccgagatttctctctcttcttcttcgtctattcttaaccgaa 245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1494 GAGNTTTATTATNNNNTAGAGTAGTATTTTTTATAGTTCTCTTAAGTAATTAATNTTAGNTT 1435

Qy 246 accaatcttatattcttccaattagaaactgaaattatttaattataataaaactgaa- 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1434 AGTTATTATAATTAATAATATTATAATAAAGTTTAGTAANNAGTTTNAAGTGCAANTATAT 1375

Qy 305 -aacaattgggtacaattcatatcacatgccttagtaataaaatgcgataaatttgata 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1374 TTATATATTTTAAATTAATTAATTAATNTTTAGTATTANTAAGATTTAAGTTATAGTTA 1315

Qy 364 aatctgcaaaagattttacaataatctcttcagaaaaaaattataacaagaatttgcgttt 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1314 GATAAATACTATNTNTTAAATATATTATNAGTTTTTAGTAGNTATTTTTTATATATTTTAAAGT 1255

Qy 424 tcatgggttggtctgaggaggatttggccactatagaactctcctcaocggaccactcttg 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1254 TAAATATTGNTNANNVGGNTAGTATATCTTATTATTATAGTATAATAATATATNTAGGNA 1195

Qy 484 cacttcaactaaaacgattgctcagaattggtgggattttatat--tcagacatatccctt 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1194 AGNTGCTAGCTTANGTAGTTAAATATTTTAAAAATAAAAAATTAATTAAGTATATCTGTN 1135

```











Qy	107	taataaaatttcacatttaaaataagaatttttttttccaaataaataataattattagtt	166
Db	1101	TAATTAGATATTTTAATAAATANTAAATATATTANACAGATATTACTTAATTATTTTAAAT	1160
Qy	167	ttattgaaataattaaattagaaattttgaaatcccgattctctoctctttcttcgccta	226
Db	1161	TTATTTTAAATATTAACTACNTAAGCTAGCAGCTTNGCTANATATATTATTATACATA	1220
Qy	227	tcatacatatttcaeccaaacca-----actttatattgtcttcaaataggaaact	276
Db	1221	AATPAGATATACTNCCANNNTANCAATAATTACTTTAAATATATAAAAATANTCTAC	1280
Qy	277	tgaattatttaattataattaaactgaaacaaatttggatcataattcatatcacatgcta	336
Db	1281	TAAAACTNAAATATATAAANAATAGTATTTATCTPAACATATAAAGCTTAAATCTTANTATA	1340
Qy	337	gtaataaaatgcgataatttaattgtataaacttcgcaaaagatttcacaaatctcttcaga	396
Db	1341	CTAANAATTTAATTAATTAATTAANAATATATAATATANTTGGACCTTNAACINTTACT	1400
Qy	397	aaaaattaataacaaattttt	416
Db	1401	AAACTTTATTATAAATATTAT	1420

RESULT	13
N60472	standard; DNA; 4590 BP.
ID	N60472;
AC	24-AUG-1991 (first entry)
DE	Sequence encoding the ring-infected Erythrocyte Surface Antigen
DE	(RESA).
DE	Malaria vaccine; antigen; epitope; ss.
KW	Malaria vaccine; antigen; epitope; ss.
OS	Plasmodium falciparum.
FT	Key
FT	Location/Qualifiers
FT	801..995
FT	exon

```

FT      W08601802-A.
PN      27-MAR-1986.
PD      PD
PF      11-SEP-1985; 006960.
PR      11-SEP-1984; AU-007067.
PR      11-SEP-1984; AU-007066.
PR      10-SEP-1985; AU-047326.
PA      (HALL-) HALL INST MED RES.
PI      Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;
PI      WPI; 86-094065/14.
PT      P-PDSB; P60569.
PT      DNA coding for Plasmodium falciparum antigens - expressing
PT      poly.peptide(s) having antigenicity of RESA or FIRA antigens of P
PS      falciparum
PS      Claim 4; Fig 1; 55pp; English.
CC      The inventors claim a novel DNA molecule which comprises a
CC      nucleotide sequence corresp. to all or a portion of the base
CC      sequence coding RESA (N60472) or FIRA (N60473). RESA and FIRA have
CC      antigenicity suitable for providing protective immunity against
CC      Plasmodium falciparum malarial infections.
SQ      Sequence 4590 BP; 1933 A; 437 C; 673 G; 1547 T;

Query Match          3.6%; Score 56; DB 1; Length 4590;
Best Local Similarity 52.1%; Pred No. 0.26; Indels 0; Gaps
Matches 125; Conservative

Qy    904   tactaatattttgggttaaataattaacattatttttaagatatattaagaataaa 963
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     464   TATTATTTTATGATAGCGCAAAAAAAAAAAAAATAAAATAATTTTTTAAAAATA 523

Qy    964   agatatttttaaaaaaatgataaaaattatatttcatcattgattttcacatcattgattc 1023
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     524   TTATATATTTAAAATATTTATATATTTATTTTATTTTATTTTATATTTTATTTATTTAT 583

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 02:38:02 ; Search time 5217.98 Seconds  
(without alignments)  
538.352 Million cell updates/sec

Title: US-09-464-528-15  
Perfect score: 1574  
Sequence: 1 atcgtatgagacatgttatt.....gttttgaagtataaagatg 1574

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pr1.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: gb\_ro.\*  
13: gb\_sts.\*  
14: gb\_sy.\*  
15: gb\_un.\*  
16: em\_fun.\*  
17: em\_hum1.\*  
18: em\_hum2.\*  
19: em\_in.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_sy.\*  
29: em\_un.\*  
30: em\_v1.\*  
31: gb\_htg1.\*  
32: gb\_htg2.\*  
33: gb\_in1.\*  
34: gb\_in2.\*  
35: em\_ba1.\*  
36: em\_ba2.\*  
37: em\_hum3.\*  
38: em\_hum4.\*  
39: gb\_pr4.\*  
40: gb\_htg3.\*  
41: gb\_htg4.\*  
42: gb\_htg5.\*  
43: gb\_htg6.\*

44: gb\_htg7.\*  
45: em\_htg1.\*  
46: em\_htg2.\*  
47: em\_htg3.\*  
48: em\_hum5.\*  
49: gb\_pl3.\*  
50: gb\_pr5.\*  
51: gb\_htg8.\*  
52: gb\_htg9.\*  
53: gb\_htg10.\*  
54: gb\_htg11.\*  
55: gb\_htg12.\*  
56: gb\_htg13.\*  
57: gb\_htg14.\*  
58: gb\_in3.\*  
59: gb\_htg15.\*  
60: gb\_htg16.\*  
61: gb\_htg17.\*  
62: em\_htg4.\*  
63: em\_htg5.\*  
64: em\_htg6.\*  
65: em\_htg7.\*  
66: em\_hum6.\*  
67: gb\_htg18.\*  
68: gb\_htg19.\*  
69: gb\_htg20.\*  
70: gb\_htg21.\*  
71: gb\_htg22.\*  
72: gb\_htg23.\*  
73: gb\_htg24.\*  
74: gb\_htg25.\*  
75: gb\_htg26.\*  
76: gb\_htg27.\*  
77: gb\_htg28.\*  
78: gb\_htg29.\*  
79: gb\_htg30.\*  
80: gb\_htg31.\*  
81: gb\_v11.\*  
82: gb\_v12.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	118.4	7.5	4601	33	DMU11584
2	118.4	7.5	19517	58	DMU37541
3	99.8	6.3	86829	33	PFMAL3P5
4	98.4	6.3	86829	33	PFMAL3P5
5	97.4	6.2	14867	34	AE001398
6	92.6	5.9	146285	39	AC005083
7	91.8	5.8	67970	33	PFMAL1P3
8	91.2	5.8	153098	33	PFMAL3P2
9	91.2	5.8	5371	7	YSCWTTGSA
10	89.6	5.7	67970	33	PFMAL1P3
11	89.6	5.7	104992	41	AC005504
12	89.6	5.7	130281	60	AC004157
13	89.6	5.7	161891	55	AC008206
14	88.8	5.6	253307	33	PFMAL3P7
15	88.2	5.6	80518	31	PFMAL13PA
16	88	5.6	170125	41	AC007465
17	87.4	5.6	4601	33	DMU37541
18	87.4	5.6	19517	58	DMU37541
19	86.8	5.5	104992	41	AC005504
20	86.8	5.5	130281	60	AC004157
21	86.8	5.5	152209	11	HS1108D11
22	86	5.5	15421	33	PFCOMP1RB
23	85.2	5.4	14001	33	PFCOMP1RB
24	85.2	5.4	15421	33	PFCOMP1RA
					U11584 Drosophila
					U37541 Drosophila
					AL034556 Plasmodiu
					AL034556 Plasmodiu
					AE001398 Plasmodiu
					AC005083 Homo sapi
					AL031746 Plasmodiu
					AL034558 Plasmodiu
					J01462 Yeast (S ce
					AL031746 Plasmodiu
					AC005504 Plasmodiu
					AC004157 Plasmodiu
					AC008206 Drosophil
					AL034559 Plasmodiu
					AL109815 Plasmodiu
					AC007465 Homo sapi
					U11584 Drosophila
					U37541 Drosophila
					AC005504 Plasmodiu
					AL034559 Plasmodiu
					AL034419 Human DNA
					X95275 P.falciparu
					X95276 P.falciparu
					X95275 P.falciparu

```

25 85 5.4 161286 67 AC025120
26 85 5.4 173915 9 AP001254 Homo sapi
27 84.8 5.4 1897 8 SC032857
28 84.8 5.4 3542 7 YSCMTG016
29 84.8 5.4 157781 55 AC011355
30 84.8 5.4 175516 60 AC006280
31 84.4 5.4 34996 34 UB7145
32 83.6 5.3 163678 52 AC010178
33 82 5.2 6591 7 YSCMTG006
34 81.6 5.2 2226 33 MIDVTRN
35 81.6 5.2 75076 39 AC004948
36 81.2 5.2 152209 11 H51108D11
37 81 5.1 14667 34 AF001398
38 80.8 5.1 1400 34 AF151382
39 80.4 5.1 85779 8 SC0011856
40 80.4 5.1 146285 39 AC005083
41 80.2 5.1 110000 31 PFMA14P1_0
42 80 5.1 94384 39 AC011718
43 79.6 5.1 161891 55 AC008206
44 79.6 5.1 193561 39 AC004801
45 79.4 5.0 170109 39 AC008079

```

## ALIGNMENTS

```

RESULT 1
DMU11584 4601 bp DNA INV 23-JUL-1994
DEFINITION Drosophila melanogaster Oregon-R mitochondrial A+T region.
ACCESSION U11584
VERSION U11584.1 GI:508826
KEYWORDS mitochondrial DNA; A+T region; tandem repeats.
SOURCE fruit fly.
ORGANISM Mitochondrion Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 4601)
Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
Sequence, Organization and Evolution of the A+T Region of
Drosophila melanogaster Mitochondrial DNA
Mol. Biol. Evol. 11, 523-538 (1994)
94285822
2 (bases 1 to 4601)
Kaguni,L.S.
Direct Submission
Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D, Dept. of
Biochemistry, Michigan State University, East Lansing, MI,
48824-1318, USA
FEATURES
source Location/Qualifiers
1. 4601
/organism="Drosophila melanogaster"
/organella="mitochondrion"
/strain="Oregon-R"
/db_xref="taxon:7227"
/dev_stage="embryo"
1. 4601
/gene="mt:ori"
/note="Mitochondrial origin"
/allele=""
/db_xref="FlyBase:FBgn0013687"
650..1022
/gene="mt:ori"
/note="repeat I-A"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
1023..1360
/gene="mt:ori"
/note="repeat I-B1"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
1361..1705

```

```

/gene="mt:ori"
/note="repeat I-C/A"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
1706..2043
/gene="mt:ori"
/note="repeat I-B2"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
2044..2388
/gene="mt:ori"
/note="repeat I-C"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
2491..2511
/gene="mt:ori"
/note="deoxythymidylate stretch"
/db_xref="FlyBase:FBgn0013687"
2512..2648
/partial
/gene="mt:ori"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
2649..3112
/gene="mt:ori"
/note="repeat II-A"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
3113..3576
/gene="mt:ori"
/note="repeat II-B1"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
3577..4040
/gene="mt:ori"
/note="repeat II-B2"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
4041..4504
/gene="mt:ori"
/note="repeat II-C"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
complement(4565..4585)
/note="deoxythymidylate stretch"
BASE COUNT 2271 a 131 c 74 g 2125 t
ORIGIN

```

```

Query Match 7.5%; Score 118.4; DB 33; Length 4601;
Best Local Similarity 48.4%; Pred. NO. 1.2e-06;
Matches 520; Conservative 0; Mismatches 531; Indels 23; Gaps 6;

Qy 104 acttaataaaattttacatttaattagaatttttttatacaataataatttata 163
||||| ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 2187 ACTTAAATAAATATTTTAAATAAATAAATATTTTAAATAAATAAATAAATA 2246
Qy 164 gttttattagaaa-tattaattagaaaattttgaatcccgattctctctcttc 222
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 2247 TTTTATATAAATATTTTATATAAATAAATAAATAAATAAATAAATAAATA 2306
Qy 223 gctattcatcttttctaaccacaaactcttatgtcttcaaatagaaactgaaat 282
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2307 TATATATATATATATATATAAATAAATAAATAAATAAATAAATAAATAAATA 2366
Qy 283 tattaattataataactgaaacaaatttggtatcaattcatatcacatgtttgata 342
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2367 TAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2423
Qy 343 aaatgcgataattatgataaaactgcaaaagattttacaaatctcttcagaaaaat 402
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2424 AAATTTAATAACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2483

```



```
trNA
171..239
/gene="mt:ND6"
/db_xref="FlyBase:FBgn0013685"
/product="trNA-Phe"
240..1265
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="NADH dehydrogenase subunit 2"
/protein_id="AAC47811.1"
/db_xref="GI:1166530"
/translation="MENNSSKILFTIMLIGTLIVTNSWLGAWGLINLLSFIPL
LSDNNMLSTASLYFLUQVLASTVLLFSSILLMLKNNNEINESFTSMIINGALL
LKSAGAPFHFPPNMEGUTWMNALMTWQIAPLMISYINIKYLLLSIVLSVII
KAIGNOTSRLKLMFASSINHLGMLSSIMSESIWLFYFLSVFLVFMENIF
KLFHNLQFSWPNVKILKFTLMNPLSLGGLPPFLGFLPKWLVIQQLTLCNQYEMLT
IMMSTLITLFFYLRYCYSAPMNNYFENNWINNMNNSINYNMYIMTFFSIFGLFI
SLFYPMF"
1264..1329
/gene="mt:ND6"
/product="trNA-Trp"
/db_xref="FlyBase:FBgn0013685"
complement(1322..1383)
/product="trNA-Cys"
complement(1403..1468)
/product="trNA-Tyr"
join(1470..1472,1474..15009)
/codon_start=1
/exception="mechanism underlying reading frame shift after
first codon uncertain"
/transl_table=5
/product="cytochrome c oxidase subunit 1"
/protein_id="AAC47812.2"
/db_xref="GI:7412849"
/translation="MSROWLSTNHKDIGTLVIFCAWAGVGTSLILIRAEGLHGP
ALIGDQIVNVTAFHAFIMFFMYPIMIGGFWNLVPLMLGAPDMPAPRNNKSF
LLPALSLLLSVSMVENGACTGWTVPPLSAGIAGGASVDLAISLHLAGISSILGA
VNFITINMRSTGSLDRMPLFVWSVITALLLSLPVLGAIITMLLTDRLNLTSE
FDPAGGDDILYOHGILFHFEGHPEVYILILPGFMISHIISQESGKKEFTGSLGIYAM
LAIGLGFVFAHMTFGMDVDTRAYFTSATMIAVPTGIFKISWLATLHGTQLSYS
PAILWALGVFTVGGTGVLANSSVDIILHDTYVVAHFYVLSMCAVPAIMAGF
IMWYPLTGLTLNNLWLSKSHFTIMRIGNVNLTFPOHFLGLACMPBRYSDYDAYTWN
IVSTGTSISLILGILFFELIWESVRSQOVYIPIQLNSSIEYQNTPPAHSYSELP
LUN"
3012..3077
/gene="mt:ND6"
/product="trNA-Leu"
/db_xref="FlyBase:FBgn0013685"
3083..3767
/note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"
/codon_start=1
/transl_except=(pos:3767,aa:TERM)
/transl_table=5
/product="cytochrome c oxidase subunit II"
/protein_id="AAC47813.1"
/db_xref="GI:1166532"
/translation="MSTWANLQDSAPLMEOQLIFFHDLALLILVMITVLVGLMFM
LFPNVYNRFLHGLIEMIWITLIPAILLILFALPSRLLYLLLDINEPSVTLKSIGH
QWYSEYSDPNNEIFSDSMIETNELMDGDFRLLDVDRVPLPMNSQIRILYITADVI
HSWTYALGVKVDGTPGRUNQTNFFINRPLGVGOCSEICGAGNHSFMPVIESVPVNY
FKWISSNNS"
3768..3838
/gene="mt:ND6"
/product="trNA-Lys"
/db_xref="FlyBase:FBgn0013685"
3840..3906
/gene="mt:ND6"
/product="trNA-Asp"
/db_xref="FlyBase:FBgn0013685"
3907..4068
/gene="mt:ND6"

CDS
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="ATPase 8"
/protein_id="AAC47814.1"
/db_xref="GI:1166533"
/translation="MPQMAPISWLLLIIFISITIFLFCISINYSYMNPSPKSNELKNI
NLNSMNWKW"
4062..4736
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="ATPase 6"
/protein_id="AAC47815.1"
/db_xref="GI:1166534"
/translation="MNTNLSVDFDPLAIFNFSNLNMLSTFLGLMIPSIYWLMPRSYNI
MNSLTLTHKEFKTLGPGSHGSTFIFISLPSLILNFMGLFPYIFTSYSLTTLT
LSLAPLKLCFMYLGMNHTOHMAHLVPOGTPAILMPMVCIEISNIIRPCTLAVR
LTANNIAGHLILLTLGNTGSSMSYMLTFLLMAQIALLVLESAYAMIQSYVAVLSTL
YSSEVN"
4736..5524
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="cytochrome c oxidase subunit III"
/protein_id="AAC47816.1"
/db_xref="GI:1166535"
/translation="MSTHSHHPFLHVDYSPWPLTGAIGAMTIVSGMKVWFHYDISLF
VLGNIITILTYQWRDVSRECTYOGHLYYAVTIGLRWGMILFISELVFFYSPFWAF
FHSLSPAELGASNPICGIIISFNFPQIPLLNLAILLASGAVTVTAHSHLMENNSQT
TOGLEFTVLLGIYFTILOAYEIEAPFTIADSIYGSTFMATGFGHGIHLIGTITFLVY
CLLRHLNNHFSKHHFGEPEAAWYHFEVDVWVWLFYITIIYMWGG"
5543..5607
/gene="mt:ND6"

trNA
Query Match 7.5%; Score 118.4; DB 58; Length 19517;
Best Local Similarity 48.4%; Pred. No. 8.2e-07;
Matches 520; Conservative 0; Mismatches 531; Indels 23; Gaps 6;

QY 104 actataaaattttacatttaatttagaattttttttatcaataataataatttatta 163
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17103 ACTATAAAATAATTTTTTTTTAAAAAAATAATTTATTAATAATTAATAAATA 17162

QY 164 gttttattagaa-tattaattagaaattttgaatcccgatttctctcctttcttc 222
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17163 TTTTATAATAAATTTATTTATAATAAATAATTTATAAATAATAATAAATAATA 17222

QY 223 gctattcatcttttctaaacaaacatttataatgtcttccaataatagaacattgaaat 282
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17223 TATATATATATATATATATAAATAAATAAATAATTTTAAATTTTAAATAATAAATA 17282

QY 283 tattaattataataacgaacaaatttggtatcaattcatatcatagcttagtaata 342
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17283 TATAATAATAAATAATAATAAATAATCATTCATTTATTAATAATTTAATAATAATA 17339

QY 343 aaatgcgataaattgataaaactctgcaaaagatttttacaataatcttccagaaaaat 402
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17340 AAATTTAATAAATAATAATAAATAAATAATTTATTTATTAATAATAATAATAATA 17399

QY 403 taatacaaaatttgcgttttcatggtggtgagagagatttgcacatagaaac 462
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17400 AAAAATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTA 17459

QY 463 tctctacagaccattcttgcacttcaactaaacatggtcagaattggtggggatttt 522
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17460 TTTATATATTTGAATATTTATTAATAATATATATATATATATATAGAAAAATTAATTTAA 17519

QY 523 atattcaagcataccctttcaaaacttctacttacttctgctgcgttcgtaataatcggtaa 582
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17520 ATAAATTTAATAAATAATTTTAAAAAATTTCT--TAAATGTATTATTTTATATAAAAAATAT 17577
```







```
Oy 661 actattttttttgtaattgacgaacaaacacagatttatcctggaatttaggaacc 720
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39502 ATAAATTTATTTTATTTAAATAAATAATATATATTTATTTATTTAAATAA 39443

Oy 721 acagatgtaacaaacaaatattatttttctaaacaaatcattgcagcagtc 780
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39442 ATATATTAATTTATTTTATTTTAAATAAATAATATATATTTATTTTAA 39383

Oy 781 ctgagccatgaaaaaacctcataaaatactcacacattgaccattgaaagttcgtt 840
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39382 TAAATAATATATATTAATTTATTTTATTTTAAATAAATAATATATTTATTTTAA 39323

Oy 841 ctcccatgggtgaacagatcaaacctcacacataacatacattggtgattcctacc 900
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39322 TTTTAAATAAATAATATATATTTATTTT-----TATTTTAAATAAATAATA 39272

Oy 901 tcatcactaatttttgggttaataattatcattatttttgaagataatttaagaat 960
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39271 TTTAATTTATTTTATTTTAAATAAATAATATATTTATTTTATTTTAAATAAATA 39212

Oy 961 taaagattttttaaaaaatgataaaattattatttattcattgatttttcacatttga 1020
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39211 TATTTATTTATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 39154

Oy 1021 ttttgaataaataatttttttcttaatttcttaaaaaatgttgcaagacacttaga 1080
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39153 TTAATATATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 39094

Oy 1081 catagctctgtctgtttcaaaagcattcatcatttaatacattaaaaattttaata 1140
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39093 TATATATACTATTACATATATATTTTATTTTATTTTATTTTAAATAATTTATTTATTT 39034

Oy 1141 ctcaacagtagaatctctctgtgagt 1165
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39033 TAAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 39009

RESULT 4
PFMAL3P5 86829 bp DNA INV 11-FEB-2000
LOCUS Plasmodium falciparum MAL3P5, complete sequence.
DEFINITION AL034556 AL008971 AL008972 AL008978 AL010141 AL010153 AL010162
ACCESSION AL010206 AL010210 AL139179
VERSION AL034556.2 GI:4493931
KEYWORDS HTG.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 86829)
AUTHORS Bowman, S., Lawson, D., Basham, D., Brown, D., Davies, R.M., Devlin, K.,
Chillingworth, T., Churcher, C.M., Craig, A., Galloway, N., Hamlin, N., Harris, D.,
Holt, R.A., Hornsby, S., Karpman, A., Kiehl, J., Krasnow, J., Kroll, M.,
Mungall, K., Murphy, L., Oliver, K., Quail, M.A., Ramesh, M.,
Ratcliffe, J., Rutter, S., Skelton, J., Squares, R., Sulston, J.E.,
Sulston, J.E., Whitehead, S., Woodward, J.R., Newbold, C. and
Barrell, B.G.
TITLE The complete nucleotide sequence of chromosome 3 of Plasmodium
falciparum
JOURNAL Nature 400 (6744), 532-538 (1999)
MEDLINE 99376085
REFERENCE 2 (bases 1 to 86829)
AUTHORS Bowman, S., Skelton, J., Churcher, C., Lawson, D., Quail, M. and
Barrell, B.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 86829)
AUTHORS Lawson, D., Bowman, S. and Barrell, B.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium,
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA, UK
COMMENT On Mar 24, 1999 this sequence version replaced gi:4034877.
For more information about this sequence or the Malaria Project,
```

```
see http://www.sanger.ac.uk/Projects/P_falciparum.
Location/Qualifiers
1. 86829
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="3"
/clone="MAL3P5"
324. 2944
/gene="MAL3P5.1"
join(324..668,1199..1303,1460..2944)
/gene="MAL3P5.1"
/note="predicted using hexExon; MAL3P5.1 (PFC0575w),
hypothetical protein, len: 645 aa"
/codon_start=1
/protein_id="CAB38969.1"
/db_xref="GI:4493933"
/db_xref="SPTREMBL:O97258"
/translation="MYLKNVYIYISCFILDFCLFSLHLLKMKYKNNHNNMKSVTFFEL
RSPQYRKRFRSRIRIKNVYFQKKQKPLFLFENLKKGFSGFLGFWNQYDQVKNRKKK
KKKKKKKKPKVHSILNOISEKVEKKDAENYALHLFLKLDENITLFSNMHIMDF
FKSKQVIECIRDIKSKRRKNLSIYNLFICTLIYFYCHCLIKVISHLICIFEFF
FFCFCLYNLIERIYECVGDILRKRIERNYNDINNDISDKMKIKFMKDAIMNNDIVNTNI
LYFNHYDELRCFTMKLNIERNKNIIRSNYNDINNDISDKMKIKFMKDAIMNNDIVNTNI
SNDEIKQFENPDENELKDYEQOLFQNDNIKYIEDQPLYNNDNSNINNDNN
NNNTMKNKKIKTDYDDDDYDEKEDLVQNDIDYIYKNTIGMKNKLEFKNQF
TEADIEFQNFSLNVLQHGKRVKSDENTKSTELKKNKNTKNGYDTELIONMENN
FKKNDNNISNDISNENIITKLAKNDSDEINLTSDLIYERURTKVLTQIKIEY
LKFKYQYDIINEQYIIRKNTVLDLLNYGKIVMSPDVDSNLSFETKIDKIDIPNEKDK
NNQENQKSNKYIIMFNV"
complement(3356..6646)
/gene="MAL3P5.2"
complement(3356..6646)
/gene="MAL3P5.2"
/note="predicted using hexExon; MAL3P5.2 (PFC0580c),
hypothetical protein, len: 1097 aa"
/codon_start=1
/protein_id="CAB38968.1"
/db_xref="GI:4493932"
/db_xref="SPTREMBL:O97257"
/translation="MLGUKRNVFYLLVSVPSLFPAYFLKRRKDNENNYETLINNDIE
KRRRIHNKCSYIPLEFNIDYIYKNNILRWLYFKRKRKKDEEYIYITNVRK
KREAIYKFISEQNLNFYIYVLEYSKYLGLSPHLSLYLKNISEHCVNIYK
SLYYNKLNDKHLNKKKLYFKQIINNEHTQAPNTHTNNNNNNKKPLDINHSC
KNTNLSYTNNEKENIYDKYNIHNFYTESYKSDENCQHTLNLVYLLQTY
DNICRCLTNTYINFYMINLYKCYKNNELIILAYNHIEDMKKKNOKNTNTS
LFYIYSSFFKKEENHIYDFEDQMHHLKKNDEKDFYNSNENHNNIYKISNY
FYDHINSNRCSFKNLKQDDNTKTHIMGKEYPMKNSDKNNNNNNTCNINEX
DOKDKILAKYFLKGNLDDIOILNELYVIMYRLLFECSLKLSIKKNIHLKKME
PDKDNKIYLSADYNNLRNLKRFKSEENRINSPFPLSLKNIYFEDEIG
RSDNTIYNNYDKETNTTNNNNNDNICSNNNDHCSNNNDHCSNNNDHCSN
NNNCSNNNNICSNNNNICSNNNDHCSNNNDHCSNNNDHCSNNNDHCSN
YELAKIYNNIDYLGKKEKQNDNTNLYIKKKFPWIFLANKIKKDTSFIEH
NNNIDKNNNIIFKKYNNFESSIYFYKDIYKRYLRYIYDNIKKKFCR
FYLNKHEINRKYKRAFHYIYFDOFIINNYHIHKKNIHKLHIHKKQCKDE
IDIVFKDLYCMINNIINNIYIYHKNVNECVRYEKAYNKILLYEYLNKENIY
YANKZKLYTLNNISNDLYPNISYKNIYQNKYKKNKKNKSFYSKNDIHLLE
LYTQIQNCCDIFSYIKYFNENKPNFLYIYELHYIYISEKKKKKAFSFISSSP
YSYDVMVNSFTSFYFFLSYLLFIFLHPDMYASYIYFFKTUTYSGLPYIYVLYNNI
MVCGPKT"
complement(13655..19355)
/gene="MAL3P5.4"
complement(join(13655..17801,18019..19355))
/gene="MAL3P5.4"
/note="predicted using hexExon; MAL3P5.4 (PFC0590c),
hypothetical protein, len: 1828 aa"
/codon_start=1
/protein_id="CAB38986.1"
/db_xref="GI:4493950"
/db_xref="SPTREMBL:O97275"
/translation="MLFYKLCILFSLIIIIHKGDISSTKLLKSGNLNNKIKKGR
RYGGYNSKMYTIEKRRYIKKQKNNNNSYFCNVYKNDVNNNTAYNYIYNNP
KNKLEYEKIKNHVKKKKKIFSLKFSQNKRNKKKYFFINETSFDHDKONIKVLF
```





```

QY 587 agactttcaaaatcatttttaacccctaaacagtaaaatttgagggacaaataatattt 646
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7863 AAAAATATTATA---TTTAAATATATAAAATGCTATATATAAAATAAATCTTTTAA 7918
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 647 tcaaatgttagactatttttttttttttttttttttttttttttttttttttttt 706
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7919 TTTTATTTTATTAATAAATATTATTATTATTATTATTATTATTATTATTATTATTAT 7978
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 707 gaatttttagaacacagatgtacaaacccaattatttttttttttttttttttttt 766
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7979 TTTTAAATGTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8038
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 767 tcatgacagatccctcagccctgaaacaccccttataaaatactcacacattgacca 826
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8039 ATAATTATAATTATAATATATTTTACATATAAAATATATATA-----TTATATATATTTAA 8093
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 827 ttgaaagtctgtccctcagggtaacacagatcaaacacacacacacacacacacac 885
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8094 TTTTATTTTATTTTAAATGAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8153
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 886 atactccttaccacatactactattttttgggttaaaataatcaatcatttttttaaga 945
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8154 AAATATTATTATATATTAAATTAATAATTAATAATATATATTATTATTAAATAACTTAATTA 8213
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 946 tataattaagaataatbaaagatttttttttttttttttttttttttttttttttttt 1005
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8214 ATTAATATTATTATTATAAAAAAATAATATAAATAATATGATATATATATAATATAATATA 8273
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1006 ttttcacacatttgatttgataaaataatatttttttttttttttttttttttttttt 1065
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8274 TATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8333
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1066 aagacacttattagacatgtctgtctgtttacaaagcattcatcatttaacatt 1125
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8334 ATATTATTATGACTCTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 8393
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1126 aaaaaatattt 1136
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8394 TAATAATATAT 8404
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 6
AC005083/c AC005083 146285 bp DNA PRI 21-DEC-1999
LOCUS Homo sapiens BAC clone CTA-281G5 from 7p15-p21, complete sequence.
DEFINITION AC005083
ACCESSION AC005083
VERSION AC005083.1 GI:4150930
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Madsen,C. and Blair,T.
TITLE The sequence of Homo sapiens BAC clone CTA-281G5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 146285)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
MO 63108, USA
REFERENCE 3 (bases 1 to 146285)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 146285)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jan 12, 1999 this sequence version replaced gi:3212908.

```

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: sapiens@wustl.wustl.edu  
----- Summary Statistics  
-----  
Center project name: H\_RG281G05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

Clone CTA-281G5 is from a release of the human BAC library C1TB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>). VECTOR: pBelOBAC11  
Selection: chloramphenicol

#### NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of CTA-281G5; actual end is at 146285 of CTA-281G5

BAC CTA-281G5 contains an E. coli transposon from 10982 to 12345 that is not represented in the submitted sequence.

FEATURES	source
repeat_region	1..146285 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7p15-p21" /clone="CTA-281G5" /clone_lib="C1TB-HS-A"
repeat_region	322..358 /rpt_family="TAAAA)n"
repeat_region	808..952 /rpt_family="MER1_type"
repeat_region	1172..1194 /rpt_family="AT_rich"
repeat_region	1264..1720 /rpt_family="LI"
repeat_region	1848..1881 /rpt_family="AT_rich"
repeat_region	1882..2171 /rpt_family="Alu"
repeat_region	3782..3804 /rpt_family="AT_rich"
repeat_region	4794..4826 /rpt_family="AT_rich"
repeat_region	5764..6034 /rpt_family="purine-rich"
repeat_region	6891..7126





KKPLVNGTYKYIDPEPSLNINNYALKNOKIGIVGSGAGKSTILLSILGLINISOG  
KITVEGRDITRYNRKGDSTIGILAOSSFFVFNWNIIRTFIDPYNFDDDEIVHALKIN  
GINLGNLDLYKHHKQDKSNYKKIOTSKVINSNDNTILLNDCIRYLSVRLYLIN  
RHKTIILIDIEPIFNUNNSVHDELNSFLGKAKSFNYIRRNHPNTVLIISHHANT  
LSCDDYIYLRKGEITRYCSYEDVKTOSELHLEMD0"  
23896..31533  
/gene="rRNA"  
/note="region containing small subunit, 5.8S and large  
subunit rRNA genes and spacer regions"  
23896..31533  
/gene="rRNA"  
complement(31966..32775)  
/gene="MALP3.04"  
complement(join(31966..32476,32675..32775))  
/note="MALP3.04, conserved hypothetical membrane protein,  
len: 203 aa, similarity: P. falciparum chromosome 2,  
PF0110W, O96126 predicted integral membrane protein (255  
aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity  
in 191 aa overlap)"  
/codon\_start=1  
/product="conserved hypothetical membrane protein,  
MALP3.04"  
/protein\_id="CAB63559.1"  
/db\_xref="GI:6594247"  
/translation="MKKSYTFINVTILLFLTLFLYNYDYFSKTKFNNNIKIDIN  
RFXRIAESEOKYPHEEDPCLTNEELIRPHNDSPVLPHEHYENIDKINELINS.  
TKTKETIKMRQNYKETDNMHNWDRFMWYKWKANLYLYKVKHLINILKDLTNP  
HDKETITTIWIKQDIEYFLNQLWQVLEKRLITLLELFYNKE"  
complement(32477..32486)  
/gene="MALP3.04"  
/note="potential splice acceptor sequence"  
complement(32669..32674)  
/gene="MALP3.04"  
/note="potential splice donor sequence, aaa/gtatat"  
36657..37343  
/gene="MALP3.05"  
join(36657..36743,36864..37343)  
/gene="MALP3.05"  
/note="MALP3.05, hypothetical protein, len: 188 aa"  
/codon\_start=1  
/product="hypothetical protein, MALP3.05"  
/protein\_id="CAB63560.1"  
/db\_xref="GI:6594248"  
/translation="MRKMSGIFFIKLLICISFICFNCMKWISYRKDLLWYSEN  
CFNYSIDRSIAEGSESSEKTKVDIPNIELLSLINVEYKMKETVGFPMNNLIN  
IANVLIHISFNINENIFSLINDSSKSPVLKFLKFGSIFPHMLNNVPKLLFDLCQ  
RNPVILILGIVILAAIYVFENFKNEC"  
36744..36749  
/gene="MALP3.05"  
/note="potential splice donor sequence, aag/gtatga"  
36854..36863  
/gene="MALP3.05"  
/note="potential splice acceptor sequence"  
complement(38049..40284)  
/gene="garp"  
complement(join(38049..39995,40210..40284))  
/gene="garp"  
/note="MALP3.06, garp, len: 673 aa, similarity: almost  
identical to GARP\_PLAFF (678 aa), fasta scores: 97.6%  
identity in 678 aa overlap"  
/codon\_start=1  
/product="hypothetical garp protein"  
/protein\_id="CAB63561.1"  
/db\_xref="GI:6594249"  
/translation="MNVLFSLNYICILFFVVCITLNFSTKCFNSGLLNKQNLKSFOS  
ITGLLNETLEKNKDDNSKSETLLKEEKEDVDPTTSDNLKNAINNNNEISSSTOP  
TNIINVDKONENSVDDKKDKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK  
DKKKEENSVSLTKQHPKPNATEHGENLYEENVNNNAQGGILLSSPYQRE  
QGGCGIISSEVHNTSDTKDNKINSEKEDHQOEEMLKTLDRKKERQKEKKEDE  
KIEKKKKKQKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK  
HDKNEETMOQDPTSETNNEINVPLSPITDVTTPPEEHKEGHEKHEGHEKKEG  
EKKEEHEKHEKHEKHEKSKGKKGKKGKKKKKKKKKKKKKKKKKKKKKKKKVKNVIE

DKCGVELINLEDKACEQHTVBSRPLSQPOCKLIDPEBOLTLMDSKVEEKNLSIQ  
EOLGTTIGRYNVVPRDNHKKMAKIEAELOKOKHYDKEDKKEEKSKEVEQ  
EDBEVEDE  
DDDE  
complement(39996..40005)  
/gene="garp"  
/note="potential splice acceptor sequence"  
complement(40204..40209)  
/gene="garp"  
/note="potential splice donor sequence, aag/gtaaca"  
45401..50233  
/gene="MALP3.07"  
Query Match 5.8%; Score 91.8; DB 33; Length 67970;  
Best Local Similarity 46.3%; Pred. No. 0.0012;  
Matches 487; Conservative 0; Mismatches 547; Indels 17; Gaps 5;  
Oy 92 aaaaatcgagacttaataaaattttacatttaaatagaaattttttttatcaataaat 151  
Db 8514 AAT 8573  
Oy 152 attaatatttagttttattagaaatattattagaaattttgattcccgatttctcc 211  
Db 8574 ATAAATTAT 8627  
Oy 212 tcttttctcgctattcatctatttcttaacaaacacatttcttcttcaaat 271  
Db 8628 ATAAT 8687  
Oy 272 gaacttgaaatttatttaataataaactgaaacaaattttggtatcaattcatatacat 331  
Db 8688 TAAATTAAT 8747  
Oy 332 gcttagtaataaaatcgataataattgataaaactgcacaaagattttacaaatatt 391  
Db 8748 GTTAAAT 8807  
Oy 392 tcagaaaaataataacaaattttgctgttcttcggtggtggtggtggtggtggtggtggtggt 451  
Db 8808 ATTAT 8867  
Oy 452 cactatagaactctctcgcgacatttttgcacttcaactcaacagatggtcgaaattg 511  
Db 8868 TAAAT 8926  
Oy 512 gtgggatttattatcagacatcccccttcaaaccttccactacttctgctgctgctgctg 571  
Db 8927 TAAAT 8986  
Oy 572 gtaatcggaacttagacttcaaaatcatttttcaacccctcaaacagataatttgaagg 631  
Db 8987 TTAAT 9044  
Oy 632 acaaaaaataatttttcaaaatttgatgactatttttttttttttttttttttttttttttt 691  
Db 9045 TTTAAT 9104  
Oy 692 aacagattattctctgaatttttaggaacacagatgtaactaaacaaatattttatttt 751  
Db 9105 AATTTAT 9164  
Oy 752 ttctaaaaacaaatttcatggtgagcagatgcccagcccatgaaaaaacctataaaaaa 811  
Db 9165 TTAAT 9224  
Oy 812 tctcacacttgaccattgaaaagttcttctcccatgggtacacagatcaaatcacatc 871  
Db 9225 TAT 9284  
Oy 872 caaacataacatgatactctcttaccatcatacataatttttgggttaaatattat 931  
Db 9285 TTAAT 9338





```

/gene="PFC0175w"
/Note="predicted using hexExon: MAL3P2.4 (PFC0175w),
Hypothetical protein, len: 594 aa; Similarity with a
number of hypothetical proteins from model organisms.
C.elegans hypothetical protein (tr:O01963) BLAST Score:
361, sum P(1) = 4.0e-47, 35% identity in 227 aa overlap."
/codon_start=1
/protein_id="CAB38990.1"
/db_xref="GI:4493881"
/translation="MNPKEQNKNQDQINVTIEYKDKMMKVIKPIINDTKNNININ
YDYNINKNYNTSDCYLMNEDGILNYLYYKFKLFOENNYNYELKIDIEKME
LAILEINLNIQIDYAYHFLKAYFNNSDLLENFNNSKVLKFLKSLHLEBDILN
NNNNIDEMIKQKQFVHCKQKQFIPLEFGDIEDYTLSCGHKYSKELCN
YLKTLANDPEDDIIIECIDLCKNLIKKNKNICEEDYEKVLTLHLYIKSK
DLKCPKPKPIYIOSVWLNNNNYICGCFHCFECLHEFHRPLLSYIKWYLENN
DDHMKHPTATYKPCNCPKPIERNKGMVVCIGSYFELCDNNKNNHGGYKCN
KYLHNSKYNKKKKKKKKDDTKYTDDEKEDDKTHDNDNINNNREKRNHL
ILNRYNFKSRFDQYAEFNIHTQLFLYFNCKNYNHLHKKMFEDALIQIKR
KILKWSYVAYFVSNKSDNOKHLEFYHQGELEKLDILQTKTEDINLTQPKNNNDND
VRDQQIQTMDIEFKNICERMENFV"
/complement(21152..24194)
/gene="PFC0180c"
/complement(join(21152..23273,23390..23604,23713..23862,
2397..24194))
/gene="PFC0180c"
/Note="predicted using hexExon: MAL3P2.5 (PFC0180c),
Hypothetical protein, len: 895 aa"
/codon_start=1
/protein_id="CAB38992.1"
/db_xref="GI:4493883"
/translation="MFRACQLKSCNCEAEEMRQSNKGVEDLYGHEIINENSLKK
IIEGPDLDRAVEIGVHTREYVAITAPQVDIVTRVEVFTIETTPKIVYEK
IKERYPIAKYKVEKREVEVYKVDIVDIOYCKLYPKVEIQEINIKRPVFKK
FVERIYEVKVEKREVEVYKIPKDKSKTKKENETAGINETNDETNE
EGRNVEYVYEHNLNGLRMILNPPVVOEGIIIEHPTTEMTSTISQSTREN
VSCFCNKVQGNRTNCSLSGSCMNPRTGELKEKTNLHNPRIEVOEPKIVKN
IEVQKHVPSVDVVPVVPVVPVVPVLPKDFVVPVPRKIIPIKWLSDY
QVCIEKPKLIDQVKNIPCDVDIKYRKWKACAVNPENLQDDVHAWMMRNH.
LAERKKYGEFYFYKSNKIHVICIALLYIFNFFYFFIFFFDEENGNK
MEENNEETKEEIEIYLEDEREFINNNENINTEIKEDSNGKEINENIKEL
NENINNEINENSGKEINNNNGKEINNNNGKEINNNNGKEINNNNGKEIHNKI
SVTKENINNYENIDNFANNOKEKVAIEEHIFVRNKSNIETKAYDIFNYSKEL
RFKDNKMLQTYNIDNLENENKTYLFENSDFENEMKEDIENKNNNTSNKYIEEO
LTASLYSPHLAMYTLQNKWIQTOTLTHELKYNDFIRASVYNNYLNQSGNLYMSDIM
RNNEFLKSNPIISFPSPGNIRNENYNNLIQNDQKQSNKYEHMNNQINIDNY
27152..28729
/gene="PFC0185w"
/27152..28729
/gene="PFC0185w"
/Note="predicted using hexExon: MAL3P2.6 (PFC0185w),
Hypothetical protein, len: 526 aa"
/codon_start=1
/protein_id="CAB38993.1"
/db_xref="GI:4493884"
/translation="MVKCNDSNQSKNTRNFCNRYNEEDRKEEGMMSQKDNENYNPND
IHNNNYNNYNTNNTSVENKFSQSNQENYENFVDCPPQNAIRILKPLIOEKIYVIMK
PEIEEKIYEPVOY IEKLVEVPHVILQEKLIHIPKVPVHERIKKCSKTIOEKIYVIMK
POIKVADKIVEPVYQVEKIEIPKIMVQERIIIPVKKIYKVEKIVEIPQELKNIDI
EKVOIEPIYEPVQKDIPTQIVDRPIHVEKIVEVPHVHIYRNIVSPQIRHPKIPV
KIDAMNHPFLPLLYQEDNENMETNKNKNTMEQNKNCNCFENLRNDRSRNSHSP
VELLHNDKSKQFYNNHNTKKTISIINPNYNNYNGQTNLQFNPNNDINNNNEHII
HNNNNIPSNVNFKSHDNTVITHNNKINNMPYENNTVLKNNYSVNPQTYTTFEFANN
IYSSPSVLSSSHEHA"
/complement(29952..31299)
/gene="PFC0190c"
/complement(29952..31299)
/gene="PFC0190c"
/Note="predicted using hexExon: MAL3P2.7 (PFC0190c), PAST

```

```

Query Match      5.8%; Score 91.2; DB 33; Length 153098;
Best Local Similarity 45.7%; Pred. NO. 0.0011;
Matches 432; Conservative 0; Mismatches 508; Indels 6; Gaps 3;

QY 115 tttacatttaaatagaaatTTTTTTTttatcaataaataattatttagttttattaga 174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136984 TTAAACATATAAAGTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 175 aatatttaattgagaaatTTTgaatcccgagatttctccctcttcttcgctattcatcat 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136924 TAAATATATATATATATATATATATATATATATATATATATATATATATATATAT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 235 ttctcaacaaacaaactcttatctctcaaatagaaactgaaattattcaataataa 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136864 TTATTTCTTTTAAATTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 295 ttaaaactgaaacaaatTTTgggtatcaattcatatcatgcttagtaataaaatgagataat 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136806 TTTATATTTCACAAAAATATATATATATATATATATATATATATATATATATATATAT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 355 taattgataaaactgcgaaagattttacaaatatctttcagaaaaaaataataacaaatt 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136746 ATAATATTAAATATT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 415 ttgtcgttttcatggtgctgagaggatttggcactatagaaactctctcagcagac 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136687 TATTATTATTAACATGTCATATATATATATATATATATATATATATATATATATATAT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 475 cattcttgcacttcaactaaacgattggtcgagaattggtgggattttattattcaagcat 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136627 TAAATATCAATATTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 535 atccctttcaaaactctcacttactctcgctcgctcgtaacggttaacattagactttc 594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136567 GAAGTATATAAGTTT TCCCATATATATAATATATATATATATATAATATATATATATAT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 595 aaaaacttttcaacccctaaacagataatttgaggagacaaaaataatttttcaaat 654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136507 TAAAAATATATTATCGGGGATATATAAATTTTATTTATAGACATATATACATTTATT 136448

QY 655 tgatagacttttttttttttttttttttttttttttttttttttttttttttttttttt 714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136447 CATAACTAATAATTATATAAATTTTATTAAGCCCTTAATAAAATTTATATATATATATA
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 715 ggaacacacagatgtaactaaacaaatattttatttttttttttttttttttttttttt 774
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136387 TATATAATATATATATATATATATATATATATATATATATATATATATATAAATTT 136328

QY 775 gcatgcctcagcccatgaaacacattataaaatatctacacattgaccattgaaag 834
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136327 ACATTCATATAATATATAAATTTTTCAAAAAATAAATAATATATATATATATATATAT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 835 ttctgttctccatgggttaaccagatcaaacctcacatccacacataacatgggatactct 894
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136267 TAATATATATATATATATATATATATATATATATATATATATATATATATAAGATTTT 136211

QY 895 taccatacatacaattattttgggttaaaataatttaatttttttttttttttttttttttt 954
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136210 TAAAAATTTTATTCGTATATTTT TCCAAAAGAAAAATTTTTTTTTTTTAAAAATTTCT 136151

QY 955 agaaattaaaagatttttttaaaaaaattgataaaattattattattatttttttttttttt 1014
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136150 ATTAATTAATAATATATATATATATATATATATATATAAATAAATAAATTTGGAATTTT 136091

QY 1015 atttgatttgataaaataatttttttttttttttttttttttttttttttttttttttt 1060
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136090 TTTTACTAAAAAATAATTTTCCACTTGTTCATTTCTTTATATAT 136045

```

RESULT 9

YSCMTTGS/c

LOCUS

YSCMTTGS

5371 bp

DNA

PLN

18-APR-1994

DEFINITION Yeast (S.cerevisiae) mitochondrial Ser-tRNA and ATPase proteolipid genes.

ACCESSION J01462 J01463

VERSION J01462.1 GI:343938

KEYWORDS - ATPase; proteolipid; transfer RNA; transfer RNA-Ser.

SOURCE Yeast (S.cerevisiae) mitochondrial DNA; cytoplasmic petite mutants (rho-) ds400/a3 [1], ds401 [2].

ORGANISM Mitochondrion Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1026 to 2806)

AUTHORS Macino,G. and Trzagoloff,A.

TITLE Assembly of the mitochondrial membrane system: The DNA sequence of a mitochondrial ATPase gene in Saccharomyces cerevisiae

JOURNAL J. Biol. Chem. 254, 4617-4623 (1979)

MEDLINE 79173209

REFERENCE 2 (bases 1 to 5371)

AUTHORS Trzagoloff,A., Nobrega,M., Akai,A. and Macino,G.

TITLE Assembly of the mitochondrial membrane system. Organization of yeast mitochondrial DNA in the olil region

JOURNAL Curr. Genet. 2, 149-157 (1980)

COMMENT [2] notes that even though there are other possible coding sequences in the ds401 segment, none are sufficiently long to code for a gene product of the size of the varl protein. a model for the mitochondrial RNA processing is proposed.

FEATURES

source 1..5371

Location/Qualifiers

1..5371

/organism="Saccharomyces cerevisiae"

/organelle="mitochondrion"

/db\_xref="taxon:4932"

gene 1..5371

/gene="OL11"

/db\_xref="SGD:S0007274"

1372..1602

/gene="OL11"

/note="ATPase proteolipid"

/codon\_start=1

/db\_xref="SGD:S0007274"

/transl\_table=3

/protein\_id="AAA32169.1"

/db\_xref="GI:343938"

/translation="MQLVLAQYKIGAGISTIGLGGAGIAIVFAALINGVSRNPSIK

DTVPFMAIFGFALSEATGLFCLMVSFLLFGV"

2856..2942

/gene="OL11"

/note="codon recognized: UCA; Ser-tRNA"

/product="tRNA-Ser"

/db\_xref="SGD:S0007274"

2294 a 352 c 368 g 2191 t 166 others

BASE COUNT 2294 a 352 c 368 g 2191 t 166 others

ORIGIN 82 bp 5' to SstII site, at wild-type map units 79.1.

Query Match 5.8%; Score 91; DB 7; Length 5371;

Best Local Similarity 45.7%; Pred. No. 0.0032;

Matches 478; Conservative 0; Mismatches 550; Indels 17; Gaps 4;

Qy 16 ttattcaaaacccaataatgatgagcgaatggtgtgtgagcagatctgtttat 75

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4287 TTATCTATATCATTTAACTAATATATATTACTAAAATATCATCTATTAAATAATATAT 4228

Qy 76 tatgatttcaggcgcaaaatgcgagctacttaataaaattttacatttaaatagaatt 135

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4227 GATAATTTAATAGGTCATAGTACTTTTATTATAATAATACTTAAATATTACAT 4168

Qy 136 tttttatcaataataattattattagttttatttagaataatttaataatagaattttg 195

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4167 AAATATTATTATATTATTATTATAATATTATTATTATTATTATTATTATTATAAA 4108

Qy 196 aatccccgattctcctctcttcgtctatcatcatcttcttaacaaacaaacttta 255

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4107 TAATATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATAAA 4048

Qy 256 tatgttcttcaaatagaacttgaaattattattataattataataactgaaacaaatttgg 315

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4047 TATTAAATTTTAAACTGCTTAAATTAATAAGTTTACTTATAATAATTTTATAATAT 3988

Qy 316 atcaatcat--atcacgtcttagtaataaaatcgataaataatgataaactcgcaaa 373

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3987 TTATATTATTAAATTTTAAAGTTTATTATTATAATAATTTTATAATTAATTTATCAT 3928

Qy 374 agattttacaataatcttccagaaaaataataa--caaatgttgcgttttcattggtg 431

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3927 TTATAGTATTATTAAATTTCTATATTTTATTATAATAATTTTGTAGATTTCATATAT 3868

Qy 432 ttggtctgagagatttggcactatagaactcctcctcagcagcattcttgcacttcaa 491

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3867 TTATATTATTAAATTTGTGTAATTTATTATTATTATTCGCCGGGACCAATTCGGGTGAA 3808

Qy 492 ctacaagatggtcagaattggtggggattttctatctcaagcatatccctttcaaaacttc 551

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3807 CAACCGGATTGGCCCGCGGGTATTATTATTATTATTATTATTATTATTATTATTATAT 3748

Qy 552 ctacttacttcgttcgttcgtaacgtaacatttagactttcaaaatcatttttaacccc 611

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3747 TTATTTTATTATTATTAT---ATCTAATAATATATTCTTTAATAATAATTTCTATTATA 3692

Qy 612 ctacaacagtaatttgaaggacaaataatttttcaaaatttgatagactatttttt 671

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3691 TTAATGAATCTCAGATATTATAATAATAATTTATTATTATTATTATTATTATTATA 3632

Qy 672 ttgttaatttgacgaacccaaa-----ccagatttctcctgaatttttaggaaccac 722

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3631 TTGATAAAATTTATTATTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 3572

Qy 723 agatgtaactaaacaaatatttatttttcttcaaaacaaatcttcagcagcagct 782

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3571 TTATATATATATTATAATTAATAATAATAATAATAATAATAATAATAATAATAATA 3512

Qy 783 cagcccatgaaaaaaccttataaaatctacacattgacattgaaagttcgttct 842

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3511 TGCATCTTATTATAATATTATTATTATTATTATAATAATAATAATAATAATAATAATA 3452

Qy 843 cccatgggtacccagatcaaacctcacatccaaacatggaatctccttccatcaatc 902

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3451 CAGTTATTATTATCTATATCTTATAATAATAATAATAATAATAATAATAATAATAATA 3392

Qy 903 atactaatatttgggttaaatatcaatcatttcttcaagataatttaataagaata 962

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3391 TTAATAATTTATATGATAATTTATAAAATAATAATAATAATAATAATAATAATAATAA 3332

Qy 963 aaagatttttcaaaaaatgataaaattattattctcatgatttttcatcattgatt 1022

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3331 AAATATTATAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 3272

Qy 1023 ttgataataaatatatttttttttaa 1047

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3271 ATATATATATATATAATAATAATAA 3247

RESULT 10

PFMALIP3/c

LOCUS PFMALIP3 67970 bp DNA INV 15-DEC-1999

DEFINITION Plasmodium falciparum MALLIP3, complete sequence.

ACCESSION AL0311746

VERSION AL0311746.9 GI:6594243

KEYWORDS HTG.

SOURCE malaria parasite P. falciparum.

ORGANISM Plasmodium falciparum

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 67970)

AUTHORS Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M. and Barrell,B.

TITLE Direct Submission

JOURNAL Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

## COMMENT

On Dec 16, 1999 this sequence version replaced gi:5763807.  
For more information about this sequence or the Malaria Project,  
see <http://www.sanger.ac.uk/Projects/P-falciparum>. IMPORTANT: This  
sequence is unfinished and does not necessarily represent the  
correct sequence. Work on the sequence is in progress and the  
release of this data is based on the understanding that the  
sequence may change as work continues. The sequence may be  
contaminated with foreign sequence from E.coli, yeast, vector,  
phage etc.

## FEATURES

```

source      Location/Qualifiers
            1..67970
            /organism="Plasmodium falciparum"
            /strain="3D7"
            /db_xref="taxon:5833"
            /chromosome="11"
            complement(1748..3276)
            /gene="MALIP3.01"
            complement(join(1748..2598,2748..2848,2990..3276))
            /gene="MALIP3.01"
            /note="MALIP3.01, conserved hypothetical protein, len: 412
            aa, similarity: upF0006 family eg to
            YBL055C/YBL0512/YBL0511, YBF5_YEAST (418 aa), fasta
            scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa
            overlap)"
            /codon_start=1
            /product="conserved hypothetical protein, UPF0006 family"
            /protein_id="CAB63556.1"
            /db_xref="GI:5594244"
            /translation="MKLVFHYIKYINVLVYISIFLKSNSUKIYNDLRYISTVKNKYV
            LOIKRNLKNNHNRIMEDNESSFIDGSLTDMFDGVYNSKKHENDLQNLNRAK
            NNNVDKIIITCTLAIEDKSLKICETDPEGKFLYLSAGVHTNCFEYIDKNKHEEKE
            IIAKVEEFTKYFKNQVENSWMGNKKICDCEKDMNNLEILLENLDTIPGEKY
            NEKDEYENLKNKIIKPNRIVCIGETGLDPLRCSKYIQIKYIFQKLQVOMEN
            LPMLEHRNGSETFEKVIDYKFLERNGVHSFTDIEDIVHIVQNKRLYGVNG
            CSLKLEINAVKPIPLNLLLETDAPCGWKTHASVEYIKDTYERAYTNLKKIKN
            IIKDDNTIERNEPYNIA"
            complement(2599..2610)
            /gene="MALIP3.01"
            /note="potential splice acceptor sequence"
            complement(2742..2747)
            /gene="MALIP3.01"
            /note="potential splice donor sequence, atg/gttaa"
            complement(2849..2861)
            /gene="MALIP3.01"
            /note="potential splice acceptor sequence"
            complement(2984..2989)
            /gene="MALIP3.01"
            /note="potential splice donor sequence, aaa/gtaaaa"
            5005..5496
            /gene="MALIP3.02"
            5005..5496
            /gene="MALIP3.02"
            /note="MALIP3.02"
            /note="MALIP3.02, hypothetical protein, len: 163 aa,
            contains possible signal sequence"
            /codon_start=1
            /product="hypothetical protein, MALIP3.02"
            /protein_id="CAB63557.1"
            /db_xref="GI:5594245"
            /translation="MKLNNRVVLCPIILFFELNSVLGNRRNNINFHETNAAK
            AMRKLGSINSLKDRGDLKIKLNDKHKDSTKWKSYFISNLEEKYSOTDLFR
            KQGINEANTKIIEDROEQEYILNDEIENATREVLNENFDELVIQSFQSLIDIOS
            LNN"
            8020..10389
            /note="possible cen1, region of very high [A+T] content"
            14884..20352
            /gene="MALIP3.03"
            14884..20352
            /note="MALIP3.03"
            /note="MALIP3.03, putative ABC transporter, len: 1822 aa"
            /codon_start=1
            /product="putative ABC transporter"
            /protein_id="CAB63558.1"
            /db_xref="GI:5594246"
            /translation="MTTYKENVGISNKGKKKKSCQNISFLNLSFDWIRPLINDLIK
            GDIELNPNCRNPDVYASKEENLROIDEVEDEFYSEKNSNEHVLHHCNSDASE
            KKYVYVYHNLWSILATKFRRIILISFYILETLVTLGGKFDYDMYLEGOKIPV
            YISLDFKVPFSGVLVYVNMFFLFFALLHFLHFLHFLHFLHFLHFLHFLHFLHFLH
            NHLQNPAYNTYRKSQTEIDISRDOLFISIGNASSSSSSGKNNKNNKINDNFVEN
            DIINFIKSTKMEKDSLENRNPVNIYINMFSDVPSVTFYFVSINLPVFKIF
            MSFYVFIKIGSNSVGTALWLSIALYSAMILFPLSKFYLIYRDKRDNNHHVL
            KFLKLMFNWSEFAFYINFRKMKYCKIRLYLSNIGVFISSIDIVVYVFI
            YKDLANKKEIKETSIIIMPLYVVKILISNVANPNLVNVMGIVNIKRLNNINDH
            LYNDIKNFMRYRNYEDYNIYVDKTFLONETISDHDGTSNHLKHLKNVKKLNTN
            MFKYFFYHKNYHNKILNKQILSGLLNVDNDTKKICFQEHKSNSTYNTNSHIE
            KKEYENIHNSSTMSNEFKKKNNYIIKLENCSGFLSYDNKCDNDHLKNINEN
            LKNSMLAIIGNVGSGSAFPHSILGDFNMTHGNLYIENFFKMPILYVPQNSWLFMG
            NIRSMILFNGEYNPLYKYITLOSELLNDLSTIEHGMKYINDDHLNLSKGVKRICLA
            RALYEHVIMHKLCTDEKLIQNEILDKDLNNKNTSSYNNKSKLVNYPENEN
            YLOKCLMDNNFYLYLLDDIFSDPSISKKIFSNLCKEDNISFKGNSCFIISMKNS
            TLDNLEDILDNVOYEIVNFEIOQKLIKTRGNISEYMEKNLNNITESHGYSNLNT
            IDYTRIKLDEVELNHHKSNMIYKAYFVGNTESVSFSEIDSEIKYIKMKKKNY
            KKEHNNKNNKNNNNKDDHINIMNDNHRNYNDINLGNSTDDSPVSLGNE
            YTLDTYTSNDSKDEIVKPLKYTHHEFNKSSMPFVKSSNMNPSNFYKNEHSD
            FKGSISLETLYWFOOVGVLLTSVWIFMLISIFTDEIKFVFLTMSTISKNKSHSD
            TILQKQVRYLEYFVILPIISLVTSGICESMIYIGNITSAIKVHNHILYSLNAPLYIF
            YNNLGNLIINRFIDISAFDGYFLKRIKAFIFRCLISLLIYIWRDGFIFPFPV
            ILIYFFVFRKRSRGCQAORLYLSCHTPLCNIYSNALSGKNIINIKKNTYHLDVYE
            HYNNFRISYFFKWLINWASLYIKIFILLTTYIMHPLYASGIILKYEKRYVRI
            LSTGYCISFSARGVITKFLCDYTHIEKEMCCVQRLEPAKISNKENASMKNEL
            NVITQYKKNENISDKISAIVEYKNVSLSSIINSODDEKKYGGIKPENYVSVK
            KKITVNGTYKIDEEPSLKNIMYALKNQKIGIVGSGACKSTILLSLGLINISQK
            KITVEGRDIRYNRKGEDSIIGIAQSFVFNWNRITFDYNNFTDDELIVHAKLQK
            GINGLNDLYKMHQDKMSKKYIQTTSKYINSONDNTILLTNDICIRYLSVLRYLN
            RHKYKILIDIEPIFNLSNVDELNGLFGLGAKSFNYIIRNFPNNTVLIIISHANT
            LSCDDYIVLVRKGEITVRCSDYEDVKTSLSLHLEMD"
            23896..31533
            /gene="rRNA"
            /note="region containing small subunit, 5.8S and large
            subunit rRNA genes and spacer regions"
            23896..31533
            /gene="rRNA"
            complement(31966..32775)
            /gene="MALIP3.04"
            complement(join(31966..32476,32675..32775))
            /gene="MALIP3.04"
            /note="MALIP3.04, conserved hypothetical membrane protein,
            len: 203 aa, similarity: P. falciparum chromosome 2,
            PF0110W, O96126 predicted integral membrane protein (255
            aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity
            in 191 aa overlap)"
            /codon_start=1
            /product="conserved hypothetical membrane protein,
            MALIP3.04"
            /protein_id="CAB63559.1"
            /db_xref="GI:5594247"
            /translation="MKKSYTINVTILLFLTLFLTYNYDTSKTFNNNKKIDIN
            RFRRIAEASEEQYPMWEDFCLINEELRPHENDSPYLPENHYENIKDELINSINS
            TKIKWETIKKMRQYKQVETNDNNHNRDMHMYKWKANIYLYKVKHLNITLKDLTNPI
            HDKETITWIKWQEDIEYFLNLOVEWLRILLELFYKKE"
            complement(32477..32486)
            /gene="MALIP3.04"
            /note="potential splice acceptor sequence"
            complement(32669..32674)
            /gene="MALIP3.04"
            /note="potential splice donor sequence, aaa/gtatat"
            36657..37343
            /gene="MALIP3.05"
            join(36657..36743,36864..37343)
            /gene="MALIP3.05"
            /note="MALIP3.05, hypothetical protein, len: 188 aa"
            /codon_start=1
            /product="hypothetical protein, MALIP3.05"
            /protein_id="CAB63560.1"
            /db_xref="GI:5594248"
            /translation="MRKMNSGIFFIKLLICISFCIFECFNKCMISYRKOLLWYSEN
            CFNYSIDRSLAEGSSSEKTKVKDIPNTELLKSLNINYEYEKKNKEIVGSPMDNNLN

```

Query Match	5.7%	Score 89.6	DB 33	Length 67970
Best Local Similarity	47.3%	Pred. No. 0.0023		
Matches 506	Conservative 0	Mismatches 544	Indels 20	Gaps 7
Qy 93	aaaxgcgagcttaataaataatttcacatttaaataggaaattttttttatcaataaata 152			
Db 10119	AAATTAAATATATGAAANAATAATTTAATTTTATTAAATAATATATTAATTAATA 10060			
Qy 153	ttaattattagttttattagaaatatttaattagaaaaattttgaaatcccgattctct 212			
Db 10059	TAAAAATTAATATTAAATAAGTAAAAATATATTAAATTTTATTAAATAA--ATATAAAT 10003			
Qy 213	cccttcttcgcattcatcatcttctcaacaaaccaatcttatgtctctcaaatag 272			
Db 10002	ATTTAAATATAAATTTAAATTTTAAAAAGAAATTAATAATATATTAATTAATTAATA 9943			
Qy 273	aacttgaaattattaattataataactgaaacaaatttgggtacatcatatatacatg 332			
Db 9942	TATTAAATTTAAATTAATTTAGTTTAAATAAGTATATTAAATTA--AATAATAAATTTT 9885			
Qy 333	cttagtaataaaatggagataattgaattgataaactgcaaaagatttttcaaatatcttt 392			
Db 9884	ATTATTATAAATAAANAATAATATATT-----ATTATAATATATTAAATTAATATT 9830			
Qy 393	cagaaaaatttaataacaaattttgtcgttttcattggtgtgctgagaggatttggc 452			
Db 9829	ATTTTAAATAATAAATAATTAATAATAAATGGATATAATAATAATGATATAATAACATT 9770			
Qy 453	actatagaaactctcctacggaccaattcttgcacttcaactaaacgagtgtcagaattgg 512			
Db 9769	ANTAAATAAATATATTATTTTAAAGTAATTAATTAATATATTATATTATTTTAAATAATTATAA 9710			
Qy 513	tggggatttttatattcaagcatatccccctttcaaaccttccacttacttcgtcgctcgg 572			

LINE	TEXT	COMMENT
3304	On Apr 2, 1999 this sequence version replaced gi.4337172.	
3305	* NOTE: This is a 'working draft' sequence. It currently	
3306	* consists of 3 contigs. The true order of the pieces	
3307	* is not known and their order in this sequence record is	
3308	* arbitrary. Gaps between the contigs are represented as	
3309	* runs of N, but the exact sizes of the gaps are unknown.	
3310	* This record will be updated with the finished sequence	

Qy 803 ataaaaatatcacattgaccattgaaaagtctggtctcccatgggtaaccagatcaa 862

Query Match	5.7%;	Score 89.
Best Local Similarity	47.38;	Pred. No.

Query Match	5.7%;	Score 89.
Best Local Similarity	47.38;	Pred. No.

Query Match	5.7%;	Score 89.
Best Local Similarity	47.38;	Pred. No.





```
* 5492 6373: contig of 882 bp in length.
* 6374 6453: gap of unknown length
* 6454 7449: contig of 996 bp in length
* 7450 7529: gap of unknown length
* 7530 8139: contig of 610 bp in length
* 8140 8219: gap of unknown length
* 8220 9315: contig of 1096 bp in length
* 9316 9395: gap of unknown length
* 9396 10239: contig of 844 bp in length
* 10240 10319: gap of unknown length
* 10320 10964: contig of 645 bp in length
* 10965 11044: gap of unknown length
* 11045 11648: contig of 604 bp in length
* 11649 11728: gap of unknown length
* 11729 12695: contig of 968 bp in length
* 12697 12776: gap of unknown length
* 12777 13976: contig of 1200 bp in length
* 13977 14056: gap of unknown length
* 14057 15045: contig of 989 bp in length
* 15046 15125: gap of unknown length
* 15126 15969: contig of 844 bp in length
* 15970 16049: gap of unknown length
* 16050 16859: contig of 810 bp in length
* 16860 16939: gap of unknown length
* 16940 17662: contig of 723 bp in length
* 17663 17743: gap of unknown length
* 17744 18767: contig of 1025 bp in length
* 18768 18847: gap of unknown length
* 18848 19809: contig of 962 bp in length
* 19810 19889: gap of unknown length
* 19890 21046: contig of 1157 bp in length
* 21047 21126: gap of unknown length
* 21127 21908: contig of 700 bp in length
* 21909 23735: contig of 1829 bp in length
* 23736 23815: gap of unknown length
* 23816 25556: contig of 1741 bp in length
* 25557 25636: gap of unknown length
* 25637 26792: contig of 1156 bp in length
* 26793 26872: gap of unknown length
* 26873 28359: contig of 1487 bp in length
* 28360 28439: gap of unknown length
* 28440 29898: contig of 1459 bp in length
* 29899 29978: gap of unknown length
* 29979 31836: contig of 1858 bp in length
* 31837 31916: gap of unknown length
* 31917 33347: contig of 1431 bp in length
* 33348 33427: gap of unknown length
* 33428 34568: contig of 1141 bp in length
* 34569 35754: contig of 1106 bp in length
* 35755 35834: gap of unknown length
* 35835 37813: contig of 1981 bp in length
* 37816 37895: gap of unknown length
* 37896 39641: contig of 1746 bp in length
* 39642 39721: gap of unknown length
* 39722 41135: contig of 1414 bp in length
* 41136 42477: contig of 1262 bp in length
* 42478 42557: gap of unknown length
* 42558 44299: contig of 1672 bp in length
* 44300 44309: gap of unknown length
* 44310 45922: contig of 1613 bp in length
* 45923 46002: gap of unknown length
* 46003 47999: contig of 1997 bp in length
* 48000 48079: gap of unknown length
* 48080 49882: contig of 1903 bp in length
* 49883 50062: gap of unknown length
* 50063 51360: contig of 1298 bp in length
* 51361 51460: gap of unknown length
* 51461 53101: contig of 1661 bp in length
* 53102 53181: gap of unknown length
* 53182 54926: contig of 1745 bp in length
```

```
* 54927 55006: gap of unknown length
* 55007 56937: contig of 1931 bp in length
* 56938 57017: gap of unknown length
* 57018 57606: contig of 589 bp in length
* 57607 57686: gap of unknown length
* 57687 58632: contig of 946 bp in length
* 58633 58712: gap of unknown length
* 58713 60613: contig of 1901 bp in length
* 60614 60693: gap of unknown length
* 60694 62727: contig of 2034 bp in length
* 62728 62807: gap of unknown length
* 62808 65311: contig of 2504 bp in length
* 65312 65391: gap of unknown length
* 65392 66685: contig of 1294 bp in length
* 66686 68830: contig of 2065 bp in length
* 68831 68910: gap of unknown length
* 68911 71103: contig of 2193 bp in length
* 71104 71183: gap of unknown length
* 71184 72193: contig of 1010 bp in length
* 72194 72273: gap of unknown length
* 72274 74138: contig of 1865 bp in length
* 74139 74218: gap of unknown length
* 74219 76237: contig of 2018 bp in length
* 76238 76317: gap of unknown length
* 76318 77913: contig of 1597 bp in length
* 77914 77993: gap of unknown length
* 77994 80808: contig of 2815 bp in length
* 80809 80888: gap of unknown length
* 80889 82776: contig of 1888 bp in length
* 82777 82856: gap of unknown length
* 82857 85682: contig of 2826 bp in length
* 85683 85762: gap of unknown length
* 85763 89308: contig of 3546 bp in length
* 89309 89388: gap of unknown length
* 89389 94079: gap of unknown length
* 94080 109467: contig of 15387 bp in length
* 109468 109546: gap of unknown length
* 109547 110183: contig of 637 bp in length
* 110184 110263: gap of unknown length
* 110264 110842: contig of 579 bp in length
* 110843 110922: gap of unknown length
* 110923 111636: contig of 714 bp in length
* 111637 111716: gap of unknown length
* 111717 112030: contig of 314 bp in length
* 112031 112110: gap of unknown length
* 112111 112674: contig of 564 bp in length
* 112675 112754: gap of unknown length
* 112755 113301: contig of 547 bp in length
* 113302 113381: gap of unknown length
* 113382 113979: contig of 598 bp in length
* 113980 114059: gap of unknown length
* 114060 114698: contig of 639 bp in length
* 114699 114778: gap of unknown length
* 114779 115187: contig of 409 bp in length
* 115188 115267: gap of unknown length
* 115268 115938: contig of 671 bp in length
* 115939 116018: gap of unknown length
* 116019 116539: contig of 521 bp in length
* 116540 116619: gap of unknown length
```

Query Match 5.7%; Score 89; DB 55; Length 161891;  
Best Local Similarity 41.2%; Pred. No. 0.0021;  
Matches 432; Conservative 0; Mismatches 615; Indels 2; Gaps 2;

```
Qy 95 aatgcgactacttaaaaaattttacatttaagaattttttttatcaataaatt 154
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139941 AATGTATATATATATATATATATATATATATATATATATATATATA 139882
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 155 aatttattagttttattagaataatttaattagaataattttgaatcccgatttcctcc 214
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139881 TATATAATATATATATGAATATTAGATTATAAATGAATATAATATAATA 139822
```









Db 52687 TTATTCAATTTTGAATGTCATATTTTATATTTCAAGAAAAAATAAATAATATATAT 52746  
QY 884 gaatactctctaccactatactaatatttttgggttaaatatttaatttttttaa 943  
Db 52747 ATATATATATATATTTTATTTTATATGTTTTTATTTAAGAAATTAAGTTTTTATAA 52806  
QY 944 gatattaattaagaataataagaattttttaaaaaa---atgtataaaatttatatttc 1000  
Db 52807 ATTCCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 52866  
QY 1001 atgatttttcacacatttgatttttgataataataatatttttttgaattcttataaaat 1060  
Db 52867 TATGTTTTCCCTAAATTTGAAAAATTAAGATTAATTTATTTTATTTTATTTTATTTTATTTT 52926  
QY 1061 gtgcaagacacttattagacatagctctgtctgtttacaaaagcattcatcattaat 1120  
Db 52927 ATAAACAT 52986  
QY 1121 acattataaaatatt 1135  
Db 52987 ATATTAATTATTATT 53001

Search completed: September 2, 2000, 02:55:46  
Job time: 12210 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 00:25:15 : Search time 2111.72 Seconds  
(without alignments)  
1501.462 Million cell updates/sec

Title: US-09-464-528-16  
Perfect score: 719  
Sequence: 1 agatacaactcacatcaaaa.....gttttgaagtataaagatg 719

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*

44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: em\_est35:\*  
83: em\_est36:\*  
84: em\_est37:\*  
85: gb\_est48:\*  
86: gb\_est49:\*  
87: gb\_est50:\*  
88: gb\_est51:\*  
89: gb\_est52:\*  
90: gb\_est53:\*  
91: gb\_est54:\*  
92: gb\_est55:\*  
93: gb\_gss1:\*  
94: gb\_gss2:\*  
95: gb\_gss3:\*  
96: gb\_gss4:\*  
97: em\_gss1:\*  
98: em\_gss2:\*  
99: em\_gss3:\*  
100: em\_gss4:\*  
101: gb\_gss5:\*  
102: gb\_gss6:\*  
103: gb\_gss7:\*  
104: gb\_gss8:\*  
105: gb\_gss9:\*  
106: em\_gss5:\*  
107: em\_gss6:\*  
108: em\_gss7:\*  
109: em\_gss8:\*  
110: em\_gss9:\*  
111: em\_gss10:\*  
112: em\_gss11:\*  
113: gb\_gss10:\*  
114: gb\_gss11:\*  
115: em\_gss12:\*  
116: gb\_gss12:\*

117: gb-gss13:.\*  
118: gb-gss14:.\*  
119: gb-gss15:.\*  
120: gb-gss16:.\*  
121: gb-gss17:.\*  
122: gb-gss18:.\*  
123: gb-gss19:.\*  
124: em-gss13:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	83.6	11.6	843	122	CNS00CS1	AL059666 Drosophila
C 2	82.8	11.5	1101	123	CNS016LI	AL106896 Drosophila
C 3	81.2	11.3	1201	123	CNS016LI	AL106627 Drosophila
C 4	81	11.3	1043	123	CNS0145P	AL103735 Drosophila
C 5	79.8	11.1	1200	123	CNS016CO	AL106578 Drosophila
C 6	78.8	11.0	928	122	CNS00DKY	AL071865 Drosophila
C 7	77.6	10.8	1101	122	CNS00EVL	AL069706 Drosophila
C 8	77.2	10.7	1101	122	CNS0021J	AL061936 Drosophila
C 9	76.8	10.7	1101	122	CNS00FVG	AL071206 Drosophila
C 10	76.8	10.7	1101	122	CNS00YVL	AL069627 Drosophila
C 11	75.6	10.5	1101	122	CNS00EVL	AL069706 Drosophila
C 12	74.6	10.4	928	122	CNS00DKY	AL071865 Drosophila
C 13	74.6	10.4	1101	122	CNS001FB	AL060732 Drosophila
C 14	74.4	10.3	996	122	CNS00FVG	AL071063 Drosophila
C 15	74.4	10.3	1101	122	CNS003DQ	AL064580 Drosophila
C 16	74.2	10.3	1043	123	CNS0145P	AL103735 Drosophila
C 17	73.8	10.3	1225	123	CNS016ID	AL106171 Drosophila
C 18	73.4	10.2	987	123	CNS014FO	AL104456 Drosophila
C 19	73.2	10.2	994	123	CNS015XG	AL060732 Drosophila
C 20	72	10.2	1101	122	CNS003BD	AL064030 Drosophila
C 21	72	10.0	1101	122	CNS00EO7	AL069440 Drosophila
C 22	71.6	10.0	1101	122	CNS0022H	AL097139 Drosophila
C 23	71.4	9.9	770	113	AQ740708	AQ740708 HS_5507_A
C 24	71.4	9.9	828	113	AQ739398	AQ739398 HS_5482_B
C 25	71.4	9.9	987	123	CNS014FO	AL104456 Drosophila
C 26	71.2	9.9	820	120	B11728	B11728 T8D18-Sp6.1
C 27	71.2	9.9	867	122	CNS00CX5	AL060052 Drosophila
C 28	71.2	9.9	1201	123	CNS0167M	AL106396 Drosophila
C 29	70.6	9.8	1101	122	CNS00BOL	AL057419 Drosophila
C 30	70.2	9.8	905	122	CNS00KHX	AL077798 Drosophila
C 31	70	9.7	843	122	CNS0091L	AL053009 Drosophila
C 32	70	9.7	843	122	CNS00CS1	AL059666 Drosophila
C 33	69.6	9.7	1101	122	CNS00LVZ	AL078819 Drosophila
C 34	69.4	9.7	1101	122	CNS004ZW	AL055440 Drosophila
C 35	69	9.6	1101	122	CNS0022U	AL097152 Drosophila
C 36	68.8	9.6	1101	122	CNS00D77	AL075293 Drosophila
C 37	68.6	9.5	1187	120	B11102	B11102 F19C22-T7.1
C 38	68.6	9.5	1201	123	CNS01522	AL106121 Drosophila
C 39	68.6	9.5	1225	123	CNS0161D	AL106171 Drosophila
C 40	68.2	9.5	990	122	CNS006OI	AL065624 Drosophila
C 41	68.2	9.5	1101	122	CNS00Y2S	AL097042 Drosophila
C 42	68	9.5	734	122	CNS010MP	AL099163 Drosophila
C 43	68	9.5	1101	122	CNS000B8	AL063632 Drosophila
C 44	68	9.5	1101	122	CNS00FMC	AL070972 Drosophila
C 45	68	9.5	1101	122	CNS012BM	AL101356 Drosophila

## ALIGNMENTS

RESULT 1  
LOCUS CNS00CS1/c 843 bp DNA GSS 04-JUN-1999  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACK26H19 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCESSION AL059666  
VERSION AL059666.1 GI:4947129

KEYWORDS GSS  
SOURCE fruit fly

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 843)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osogawa and  
Aaron Mammeter in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
ECORI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source

1..843

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_lib="RPCI-98"

/clone="BACR26H19"

/note="end : TET3"

BASE COUNT 131 a 102 c 259 g 141 t 210 others

ORIGIN

Query Match 11.6%; Score 83.6; DB 122; Length 843;

Best Local Similarity 39.2%; Pred. No. 2.1e-05;

Matches 120; Conservative 64; Mismatches 122; Indels 0; Gaps 0;

QY 1 agatcaaacctcaccaaacatgacatctctccctaccatcactaactattt 60

DB 828 AWATAAWATATATWTTTATATTTTATATTTTDTWDAATAWTRTATWTTTATTT 769

QY 61 tgggttaataatcatcatttttaagatataaataaagaatttttttaa 120

DB 768 TAWTWTWTWTATTAATTAATTTATATATATWTTATATATATWTTAAATTTWTTATTT 709

QY 121 aaaaatgataaaatataattattcatgatttttcacacattgattttgataaaata 180

DB 708 AWATATATATAWATAAAATTAATTTATTTWAWMMWTTATTTATTTWTTWATAWTTW 649

QY 181 tatttttttaattcttaaaaaatgttgacacactttagacacactgtctgtctg 240

DB 648 TWTWTAWTAWTAWTAWTAWTAWTAWTAWTAWTAWTAWTAWTAWTAWTAWTAWTAWTAW 589

QY 241 tttaaaaaagcattcatcatttaataacattataaaataatttaactaacagtagaact 300

DB 588 WATAATAAAAWWATAWTAWTTTATMAAATWAAATAATAAHAAATAAAAAAAWAAATW 529

QY 301 tcttgt 306

DB 528 WATWAT 523

RESULT 2

CNS016LI/c

LOCUS

CNS016LI

1101 bp

DNA

GSS

26-JUL-1999

[illegible]

BP 191.91006 EVRY cedex - FRANCE (E-mail : [secretegenoscope.cns.fr](mailto:secretegenoscope.cns.fr))







**AUTHORS**  
**TITLE** Genoscope.  
**JOURNAL** Direct Submission  
 Submitted (23-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
**COMMENT**  
 Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the European Drosophila Genome Project (EDGP) -  
 http://www.edgp.ebi.ac.uk. This Drosophila melanogaster BAC  
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
 d'Etude du Polymorphisme Humain) with funding provided by a MRC  
 project grant. The DNA was prepared from embryos by Alain Bucheton  
 and Genevieve Payan. It has been constructed in the vector  
 pBelOBAC11.

**FEATURES**  
 source  
 1..1101  
 /organism="Drosophila melanogaster"  
 /plasmid="pBelOBAC11"  
 /db\_xref="taxon:7227"  
 /clone\_lib="DrosBAC"  
 /clone="BACN01G13"  
 /note="end : SP6"  
 BASE COUNT 424 a 119 c 129 g 248 t 181 others  
 ORIGIN

Query Match 10.7%; Score 76.8; DB 122; Length 1101;  
 Best Local Similarity 39.5%; Pred. No. 0.00029;  
 Matches 113; Conservative 55; Mismatches 118; Indels 0; Gaps 0;  
 QY 6 aaactcacatcaacataacatggatctctccatccatcatactaatattttgggt 65  
 Db 1069 ATAAATWNTATWTAAGWNNWTTATWATATTAATATATTTTWTWTTATATWTHW 1010  
 QY 66 taaatattatcatatttttaagataataattaaagaataaaagatttttcaaaaaa 125  
 Db 1009 TAAATTTTATWATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 950  
 QY 126 tgataaataatcatatcatgatttttcatacatgatttgatgataataataatttt 185  
 Db 949 WTTWTTTTTTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTT 890  
 QY 186 ttttattcttcaaaaatgtgcagacacattatagacatagctgtgtctgtttac 245  
 Db 889 TWTWTTTTTTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTT 830  
 QY 246 aaagcattcatatttaatacattaaaaataatttaataactaaca 291  
 Db 829 TWTWNTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 784

**RESULT 11**  
**CNS00EVL** 1101 bp DNA GSS 04-JUN-1999  
**LOCUS** Drosophila melanogaster genome survey sequence T7 end of BAC:  
**DEFINITION** BACR2923 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
**ACCESSION** AL069706  
**VERSION** AL069706.1 GI:4949849  
**KEYWORDS** GSS.  
**SOURCE** fruit fly.  
**ORGANISM** Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
**REFERENCE** 1 (bases 1 to 1101)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
**COMMENT**  
 Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila.

melanogaster genome using these BACs. For further information  
 please see <http://www.fruitfly.org> The BDGP Drosophila  
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPCI-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from the  
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
 p1 and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be  
 found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

**FEATURES**  
 Location/Qualifiers  
 1..1101  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="RPCI-98"  
 /clone="BACR2923"  
 /note="end : T7"  
 BASE COUNT 419 a 91 c 60 g 299 t 232 others  
 ORIGIN

Query Match 10.5%; Score 75.6; DB 122; Length 1101;  
 Best Local Similarity 36.0%; Pred. No. 0.00047;  
 Matches 141; Conservative 71; Mismatches 180; Indels 0; Gaps 0;  
 QY 14 atccaaacatacagatgatactctctccatccatcatactaatattttgggttaaatatt 73  
 Db 462 WTNMMMMMMWAAATWTAAGWNNWTTATWATTAATTAATTAATTAATTAATTTTWWWWT 521  
 QY 74 aatcattatttttaagataataattaaagaatttttcaaaaaatgtataaa 133  
 Db 522 TWTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTT 581  
 QY 134 attatatttcatgatttttcatgatttttcatgatttttcatgatttttcatgattttt 193  
 Db 582 ATATTATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTT 641  
 QY 194 tctctaaaaatgtgcagacacattatagacatagctgtctgtcttcaaaaagcat 253  
 Db 642 TTAATTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 701  
 QY 254 tcatcatttaatacattaaaaatatttaatacactagatagctgtctgtgtgaggt 313  
 Db 702 AAAAATWTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 761  
 QY 314 gtgggagtaggcacccctgcatgaaacagagagagagagagagagagagagagagagag 373  
 Db 762 ATATAWTTTWTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTT 821  
 QY 374 aaaaagtatgcaacaaacaaatcaaatcaaa 405  
 Db 822 AAATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATA 853  
**RESULT 12**  
**CNS00DKY** 928 bp DNA GSS 04-JUN-1999  
**LOCUS** Drosophila melanogaster genome survey sequence T7 end of BAC #  
**DEFINITION** BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
**ACCESSION** AL071865  
**VERSION** AL071865.1 GI:4948170  
**KEYWORDS** GSS.  
**SOURCE** fruit fly.  
**ORGANISM** Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
**REFERENCE** 1 (bases 1 to 928)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submission

**JOURNAL** Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

**COMMENT** - Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pi and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

**FEATURES** Location/Qualifiers  
source  
1..928  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACR27A24"  
/note="end : 77"

**BASE COUNT** 262 a 70 c 84 g 321 t 191 others

**ORIGIN**

Query Match 10.4%; Score 74.6; DB 122; Length 928;  
Best Local Similarity 32.8%; Pred. No. 0.0007;  
Matches 96; Conservative 91; Mismatches 103; Indels 3; Gaps 1;

QY 32 taTctctaccacatacactaattatttgggttaataataatacattatttttaagat 91  
Db 557 TWT 616  
QY 92 attaatgaagaataaagattttttaaataataataataataataataataata 151  
Db 617 WAAATATWTAT 673  
QY 152 ttctacacatttgatttgataataataataataataataataataataata 211  
Db 674 AATAAAAAATTTTAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 733  
QY 212 agacacttttagacatgcttctcttcaaaagcattcattcattcaatcat 271  
Db 734 WAAATATWT 793  
QY 272 aaaaattattactaacagtagaatcttctgtgagtggtggtgagtagg 324  
Db 794 AAAAAAAWAAAAAADDWDDWKAKKKKKKKKKKKKKKKKKKKKKKKK 846

**RESULT 13**  
CNS001FB/c 1101 bp DNA GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR04A23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

**ACCESSION** AL060732  
**VERSION** AL060732.1 GI:4939397  
**KEYWORDS** GSS.  
**SOURCE** fruit fly.  
**ORGANISM** Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

**REFERENCE** 1 (bases 1 to 1101)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

**COMMENT** - Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pi and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

**FEATURES** Location/Qualifiers  
source  
1..1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACR04A23"  
/note="end : TET3"

**BASE COUNT** 288 a 110 c 103 g 491 t 109 others

**ORIGIN**

Query Match 10.4%; Score 74.6; DB 122; Length 1101;  
Best Local Similarity 39.9%; Pred. No. 0.00069;  
Matches 163; Conservative 55; Mismatches 184; Indels 7; Gaps 1;

QY 65 ttaaatattacattcatttttaagattatttaagaaataaagattttttaaaaa 124  
Db 1000 TWAATAAAAAATTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 941  
QY 125 atgtataaataataataataataataataataataataataataataata 184  
Db 940 AAAAAAAWATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 888  
QY 185 tttttaaattttaaataatttgaagacacattattagacatttctgtctgt 244  
Db 887 ATATWAAATTAATAAAAAATTAATTAATTAATTAATTAATTAATTAATTA 828  
QY 245 caaaagcattcattcattttaaataataataataataataataataata 304  
Db 827 AAAAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 768  
QY 305 gtgagtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 364  
Db 767 WMAWAAAAAATAATTAATAWAAAAAATAATTAATAWAAAAAATAATTAATA 708  
QY 365 aagacaaataaaagatgcacaaacaaacaaacaaacaaacaaacaaacaa 424  
Db 707 AATAWATAATAWAAAAAATAATTAATAWAAAAAATAATTAATAWAAAAA 648  
QY 425 ctcaattggtgtcattcaatttttcaactcagtcacagtggtgagatt 473  
Db 647 GGGGGGGGGGGHHYHWTWHTWHTWHTWHTWHTWHTWHTWHTWHTWHTWHT 599

**RESULT 14**  
CNS00FUFH/c 996 bp DNA GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC:  
BACR31021 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

**ACCESSION** AL071063  
**VERSION** AL071063.1 GI:4951105  
**KEYWORDS** GSS.  
**SOURCE** fruit fly.  
**ORGANISM** Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

REFERENCE 1 (bases 1 to 996)  
 AUTHORS Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 TITLE Genoscope.  
 JOURNAL Direct Submission  
 COMMENT Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 - Web : www.genoscope.cns.fr  
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosier in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
 source  
 Location/Qualifiers  
 1..996  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="RPCI-98"  
 /clone="BACR31021"  
 /note="end : TET3"

BASE COUNT 383 a 164 c 81 g 171 t 197 others  
 ORIGIN

Query Match 10.38; Score 74.4; DB 122; Length 996;  
 Best Local Similarity 41.08; Pred. No. 0.00075;  
 Matches 109; Conservative 50; Mismatches 107; Indels 0; Gaps 0;  
 Qy 23 taacatggatctccctaccatcatactaatattttgggttaaatattatcattat 82  
 Db 863 TTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 804  
 Qy 83 tttaagaatataaagaatataaagaatatttttaaaaaataaataattatta 142  
 Db 803 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 744  
 Qy 143 ttcagatgttttcacattgatttgatataaataatttttttttttttttttttt 202  
 Db 743 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 684  
 Qy 203 aatgttgaagacattattagacatagcttctgttttcaaaaagcattcatcatt 262  
 Db 683 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 624  
 Qy 263 aatcattaaaaattttaacta 288  
 Db 623 WTTAKAAATTTWKAATTTWGAATW 598

RESULT 15  
 CNS003DQ/c 1101 bp DNA GSS 03-JUN-1999  
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #  
 DEFINITION BACR08109 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
 ACCESSION AL064580  
 VERSION AL064580.1 GI:4941932  
 KEYWORDS GSS.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 1101)

AUTHORS Direct Submission  
 TITLE Genoscope.  
 JOURNAL Direct Submission  
 COMMENT Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 - Web : www.genoscope.cns.fr  
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosier in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
 source  
 Location/Qualifiers  
 1..1101  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="RPCI-98"  
 /clone="BACR08109"  
 /note="end : TET3"

BASE COUNT 291 a 51 c 117 g 404 t 238 others  
 ORIGIN

Query Match 10.38; Score 74.4; DB 122; Length 1101;  
 Best Local Similarity 44.18; Pred. No. 0.00074;  
 Matches 178; Conservative 36; Mismatches 187; Indels 3; Gaps 1;  
 Qy 19 aacataacatggatctccctaccatcatactaatattttgggttaaatattatc 78  
 Db 535 MACWTAAATTTTAAATTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTT 476  
 Qy 79 ttatttcaagatataaagaatataaagaatatttttaaaaaataaataattat 138  
 Db 475 CCCWTTTTTATTTTAAATW--WAATTAATTTAAATTTAAATTTTAAATTTT 419  
 Qy 139 attattcatgttttccatcacattgatttgatataaataattttttttttttt 198  
 Db 418 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 359  
 Qy 199 aaaaaattgtcgaagacacttattagacatagcttctgttttcaaaaagcattc 258  
 Db 358 WAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 299  
 Qy 259 atttaacataaaaaataatttaataacagtagaattctcttgagtggtggg 318  
 Db 298 ATATTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 239  
 Qy 319 agtaggcaacctggcattgaacagagagagagagagagagagagagagagagag 378  
 Db 238 AATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 179  
 Qy 379 gtatgcaacacaaatcaaatcaaaagggcgaagggctgggggtt 422  
 Db 178 AA 135

Search completed: September 2, 2000, 00:25:22  
 Job time: 4346 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 02:00:43 : Search time 111.97 Seconds  
(without alignments)  
883.296 Million cell updates/sec

Title: US-09-464-528-16  
Perfect score: 719  
Sequence: 1 agatcaaacacatccaaa.....gttttgaaagtataaagatg 719

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq: \*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	7.8	19124	4	US-08-487-826B-13
2	54.2	7.5	1186	4	US-08-731-722-5
3	52.6	7.3	665	4	US-08-883-795A-1
4	52	7.2	8920	3	US-08-446-855A-1
5	51.4	7.1	19124	4	US-08-487-826B-13
6	51.2	7.1	3933	4	US-08-731-722-3
7	51.2	7.1	3933	4	US-08-731-722-3
8	51	7.1	665	4	US-08-883-795A-1
9	50.2	7.0	473	2	US-08-764-100-16
10	50.2	7.0	4970	2	US-08-764-100-14
11	50.2	7.0	4970	2	US-08-764-100-20
12	50.2	7.0	8920	3	US-08-446-855A-1
13	49.8	6.9	4467	2	US-08-565-907A-1
14	49.8	6.9	4467	4	US-08-910-551B-1
15	49.8	6.9	4467	4	US-08-909-425A-1
16	49.4	6.9	1611	7	5213972-6
17	49.2	6.8	319	1	US-07-593-657-14
18	49	6.8	4098	4	US-08-605-106-4
19	48.2	6.7	6768	2	US-08-107-755A-1
20	48.2	6.7	8457	1	US-07-991-867B-1
21	48.2	6.7	8457	4	US-08-544-332-1
22	47.8	6.6	12124	1	US-08-181-271A-36
23	47.8	6.6	12124	1	US-08-449-315-36
24	47.8	6.6	12124	1	US-08-444-803-36
25	47.8	6.6	12124	1	US-08-449-043-36
26	47.8	6.6	12124	2	US-08-456-265A-36

27 47.8 6.6 12124 2 US-08-455-416-36 Sequence 36, Appl  
28 47.8 6.6 12124 2 US-08-455-244-36 Sequence 36, Appl  
29 47.8 6.6 12124 2 US-08-454-876-36 Sequence 36, Appl  
30 47.8 6.6 12124 3 US-08-457-364-36 Sequence 36, Appl  
31 47.8 6.6 12124 3 US-08-456-262-36 Sequence 36, Appl  
32 47.8 6.6 12124 3 US-08-456-240-36 Sequence 36, Appl  
33 47.8 6.6 12124 3 US-08-455-736-36 Sequence 36, Appl  
34 47.8 6.6 12124 4 US-08-971-217-36 Sequence 36, Appl  
35 47.6 6.6 10395 1 US-08-245-809-3 Sequence 5, Appl  
36 47.6 6.6 10396 1 US-08-245-809-5 Sequence 2, Appl  
37 47.6 6.6 10798 1 US-08-107-748-2 Sequence 2, Appl  
38 47.6 6.6 10798 6 PCT-US92-01385-2 Sequence 4, Appl  
39 47.6 6.6 10965 1 US-08-107-748-4 Sequence 4, Appl  
40 47.6 6.6 10965 6 PCT-US92-01385-4 Sequence 13, Appl  
41 47.4 6.6 642 2 US-08-764-100-13 Sequence 7, Appl  
42 47.4 6.6 643 2 US-08-764-100-7 Sequence 2, Appl  
43 47.4 6.6 2993 2 US-08-764-100-2 Sequence 10, Appl  
44 47.4 6.6 2993 2 US-08-764-100-10 Sequence 9, Appl  
45 47.4 6.6 3000 2 US-08-764-100-9

## ALIGNMENTS

## RESULT 1

US-08-487-826B-13  
Sequence 13, Application US/08487826B  
Patent No. 593827

## GENERAL INFORMATION:

APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israel, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs.  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13







[illegible]

US-08-883-795A-36/c

US-08-883-733A-30/C  
: sequence 36. Application US/08883795A

Patent No. 5985607  
GENERAL INFORMATION:  
APPLICANT: Delcuve, Genevieve  
APPLICANT: Awang, Gregor  
TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERESKIN & PARR.  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/883.795A  
FILING DATE: 27-JUN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gravelle, Michelle  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 7841-062  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 665 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: Rh 32  
US-08-883-795A-36

Query Match 7.1%; Score 51; DB 4; Length 665;  
Best Local Similarity 53.0%; Pred. No. 0.035;  
Matches 131; Conservative 0; Mismatches 115; Indels 1; Gaps 1;

Qy 46 tcataactatttttggttaataatcattatttttaagatatttaataaagaat 105  
Db 508 TAAATTATAAATACTTTAATTATAAATAATGTAATTATAAATAATGTAAT 449  
Qy 106 taaagattttttaaaaaagtataaaattattattcatgatttttcacattga 165  
Db 448 TATAAATATGTAATTATAAACAATTTTAATTATAAATAATGTAATTATAAACAATT 389  
Qy 166 ttttgataataatatttttttttaatttttcttaaaagtgtgcaagacatttata 225  
Db 388 TATAAATATGTAATTATAAACAATTTTAATTATAAATAATGTAATTATAAACAATT 329  
Qy 226 catagctc-tgttcgtttacaagaagcattcatttaataacattaaaaatttata 284  
Db 328 TATAAATATGTAATTATAAACAATTTTAATTATAAATAATGTAATTATAAACAATT 269  
Qy 285 actaaca 291  
Db 268 TATAAATA 262

RESULT 9  
US-08-764-100-16  
; Sequence 16, Application US/08764100  
; Patent No. 5773700

GENERAL INFORMATION:  
APPLICANT: van Grinsven J., Martinus Q.  
APPLICANT: De Haan, Petrus T.  
APPLICANT: Gielen L., Johannes J.  
APPLICANT: Peters, Dirk  
APPLICANT: Goldbach, Robert W.  
TITLE OF INVENTION: Improvements in or Relating to Organic  
TITLE OF INVENTION: Compounds  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sandoz Agro, Inc  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764.100  
FILING DATE: 06-DEC-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214.064  
FILING DATE:  
APPLICATION NUMBER: US 08/032.235  
FILING DATE: 17-MAR-1993  
APPLICATION NUMBER: GB 9206016.9  
FILING DATE: 19-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5773700ris, Allen E.  
REGISTRATION NUMBER: 34,490  
REFERENCE/DOCKET NUMBER: 137-1061  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 354-3592  
TELEFAX: (415) 857-1125  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 473 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-764-100-16

Query Match 7.0%; Score 50.2; DB 2; Length 473;  
Best Local Similarity 52.1%; Pred. No. 0.047;  
Matches 112; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 67 aaataataatcatttttttaagatatttaataaagaataaaagattttttaaaaaat 126  
Db 236 AAAACCAAAAAAATTTTGTGTAATAATAAGGCTCGGCCAGATTGGTCTAAGACC 295  
Qy 127 gtataaaatattattcattgatttttcacattgatttttgataataataatttt 186  
Db 296 TTTTATTTGTTTTTATACATTTTATTGTTTGTGATTTTATTATTATTATTATTTA 355  
Qy 187 ttttaattcttaaaaaatgttgcaagacatttattagacattgttctctgtttaca 246  
Db 356 TATTTTTTATATAGTTTGGCTTATTTAACACTTATTTAGACAAATTAATTTTATTTA 415  
Qy 247 aaagcattcatcatttaataacataaaaaatttt 281  
Db 416 CAATCATCTGCTTATTTAATTTAAACACATTT 450

RESULT 10  
US-08-764-100-14  
; Sequence 14, Application US/08764100  
; Patent No. 5773700

## GENERAL INFORMATION:

APPLICANT: van Grinsven J., Martinus Q.  
APPLICANT: De Haan, Petrus T.  
APPLICANT: Gielen L., Johannes J.  
APPLICANT: Peters, Dirk  
APPLICANT: Goldbach, Robert W.  
TITLE OF INVENTION: Improvements in or Relating to Organic  
TITLE OF INVENTION: Compounds  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sandoz Agro, Inc  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,100  
FILING DATE: 06-DEC-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/214,064

## FILING DATE:

APPLICATION NUMBER: US 08/032,235

FILING DATE: 17-MAR-1993

APPLICATION NUMBER: GB 9206016.9

FILING DATE: 19-MAR-1992

## ATTORNEY/AGENT INFORMATION:

NAME: No. 57737001s, Allen E.

REGISTRATION NUMBER: 34,490

REFERENCE/DOCKET NUMBER: 137-1061

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 354-3592

TELEFAX: (415) 857-1125

INFORMATION FOR SEQ ID NO: 14:

## SEQUENCE CHARACTERISTICS:

LENGTH: 4970 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-764-100-14

Query Match 7.0%; Score 50.2; DB 2; Length 4970;

Best Local Similarity 52.1%; Pred. No. 0.06;

Matches 112; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 67 aaatattatcatctatttttaagatatttaataaagataataaagatttttttaaaaaaat 126

DB 1233 AAACCAAAATAATTTTTTTGTAATAATAAGGTCGCGGCAGATTGGTCTAGACC 1292

QY 127 gataaaattatattatcatgatttttttcatcattgttttgataataaataatttt 186

DB 1293 TTTTATTTGTTTTTATACATTTTATTGTTGTTGATTTTATTATTATTATTTTA 1352

QY 187 ttttaattcttaaaaaatgttgcagacactatttagacagcttctgtctgtttaca 246

DB 1353 TATTTTATATAGTTGCTTATTAAACACTATTATTAGACAAATAAATTTATTGATTA 1412

QY 247 aaagcattcatcatttaataacattaaaaaatattt 281

DB 1413 CAATCATCTGCCTTATTATTAAATTTAAACACATTT 1447

## RESULT 11

US-08-764-100-20/c

; Sequence 20, Application US/08764100

; Patent No. 5773700

## GENERAL INFORMATION:

APPLICANT: van Grinsven J., Martinus Q.  
APPLICANT: De Haan, Petrus T.  
APPLICANT: Gielen L., Johannes J.  
APPLICANT: Peters, Dirk  
APPLICANT: Goldbach, Robert W.  
TITLE OF INVENTION: Improvements in or Relating to Organic  
TITLE OF INVENTION: Compounds  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sandoz Agro, Inc  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,100  
FILING DATE: 06-DEC-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/214,064

## FILING DATE:

APPLICATION NUMBER: US 08/032,235

FILING DATE: 17-MAR-1993

APPLICATION NUMBER: GB 9206016.9

FILING DATE: 19-MAR-1992

## ATTORNEY/AGENT INFORMATION:

NAME: No. 57737001s, Allen E.

REGISTRATION NUMBER: 34,490

REFERENCE/DOCKET NUMBER: 137-1061

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 354-3592

TELEFAX: (415) 857-1125

INFORMATION FOR SEQ ID NO: 20:

## SEQUENCE CHARACTERISTICS:

LENGTH: 4970 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-764-100-20

Query Match 7.0%; Score 50.2; DB 2; Length 4970;

Best Local Similarity 52.1%; Pred. No. 0.06;

Matches 112; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 67 aaatattatcatctatttttaagatatttaataaagataataaagatttttttaaaaaaat 126

DB 3738 AAACCAAAATAATTTTTTTGTAATAATAAGGTCGCGGCAGATTGGTCTAGACC 3679

QY 127 gataaaattatattatcatgatttttttcatcattgttttgataataaataatttt 186

DB 3678 TTTTATTTGTTTTTATACATTTTATTGTTGTTGATTTTATTATTATTATTTTA 3619

QY 187 ttttaattcttaaaaaatgttgcagacactatttagacagcttctgtctgtttaca 246

DB 3618 TATTTTATATAGTTGCTTATTAAACACTATTATTAGACAAATAAATTTATTGATTA 3559

QY 247 aaagcattcatcatttaataacattaaaaaatattt 281

DB 3558 CAATCATCTGCCTTATTATTAAATTTAAACACATTT 3524

## RESULT 12

US-08-446-855A-1

; Sequence 1, Application US/08446855A

; Patent No. 5849573

APPLICANT: Kondo  
 TITLE OF INVENTION: DNA Encoding Phase  
 TITLE OF INVENTION: Abortive Infection Protein  
 TITLE OF INVENTION: From Lactococcus  
 TITLE OF INVENTION: lacticis, and Method of Use Thereof  
 NUMBER OF SEQUENCES: 1  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ian C. McLeod  
 STREET: 2190 Commons Parkway  
 CITY: Okemos  
 STATE: Michigan  
 COUNTRY: USA  
 ZIP: 48864  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette 5.25 Inch,  
 MEDIUM TYPE: 360 KB storage  
 COMPUTER: Acer  
 OPERATING SYSTEM: MS-DOS (version 4)  
 SOFTWARE: Wordperfect 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/565,907A  
 FILING DATE: December 1, 1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ian C. McLeod  
 REGISTRATION NUMBER: 20,931  
 REFERENCE/DOCKET NUMBER: Quest 4.1-152  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (517) 347-4100  
 TELEFAX: (517) 347-4103  
 TELEX: NO. 5814499e  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4467  
 TYPE: Nucleotide  
 STRANDEDNESS: Single  
 TOPOLOGY: Linear  
 MOLECULE TYPE:  
 DESCRIPTION: Genomic DNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: N/A  
 ORIGINAL SOURCE:  
 ORGANISM: Lactococcus lactis  
 STRAIN:  
 INDIVIDUAL ISOLATE: W1  
 DEVELOPMENTAL STAGE: N/A  
 HAPLOTYPE: N/A  
 TISSUE TYPE: N/A  
 CELL TYPE: Bacterium  
 CELL LINE: N/A  
 ORGANELLE: N/A  
 IMMEDIATE SOURCE:  
 LIBRARY: genomic  
 CLONE: SMG-20  
 POSITION IN GENOME: N/A  
 FEATURE:  
 NAME/KEY: phase abortive infection  
 LOCATION: N/A  
 IDENTIFICATION METHOD: sequencing  
 OTHER INFORMATION: DNA encoding phase  
 OTHER INFORMATION: resistance  
 PUBLICATION INFORMATION: N/A  
 US-08-565-907A-1

```

RESULT 13
US-08-565-907A-1/c
; Sequence 1, Application US/08565907A
; Patent No 5814499
; GENERAL INFORMATION:
; APPLICANT: Sylvain Molneau, Barbara
; APPLICANT: J. Holler, Peter A. Vandenbergh,
; APPLICANT: Ebenezer R. Vedamuthu, Jeffrey K.
; OTHER INFORMATION: resistance
; PUBLICATION INFORMATION: N/A
US-08-565-907A-1
Query Match 6.9%; Score 49.8; DB 2; Length 4467;
Best Local Similarity 49.4%; Pred. No. 0.071;
Matches 129; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

```



	Query Match	6.98;	Score 49.8;	DB 4;	Length 4467;
	Best Local Similarity	49.4%;	Pred. No.	0.071;	
Matches	129; Conservative	0;	Mismatches	132;	Indels    0; Gaps         0;
Qy	48	aatacattatcttgggttaaaataaatcaatcatattttttaagatatcaattcaagaatta	107		
Dd	2679	ATTTAAATCAATTAGITTAGTAATCTTCTTGATTAAATAAACAATCATATTTCAAAGAC	2620		
Qy	108	aagagtatttttaaasaaaatgataaaaaattattatttcatttatccacatttgatt	167		
Dd	2619	AACCATATACAACAATATTTCGATATAATCTTGATTAAAAATTATTTTTACGATAGTGTGC	2560		
Qy	168	ttagataataaatatattttttttaaattcttcaaaaaattgtcgcaagcacactattagaca	227		
Dd	2559	TATTTTTTCITTTTGAAGCCCTTTGAATATTACTAAAATATTTTTTTTACAAATATTTGAAGC	2500		
Qy	228	tagcttgttgcgttttacaaaagcattccattcataaacattaaaaaattttaatact	287		
Dd	2499	TGATPAAACTTGGAAATCCAAATTCATTAATATTTTTCAAGAAGGTCAA AAAACTATTATTAGT	2440		

GenCore version 4.5  
Copyright (c) 1993 - 2000. Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 02:07:29 ; Search time 161.21 Seconds  
(without alignments)  
1115.862 Million cell updates/sec

Title: US-09-464-528-16  
Perfect score: 719  
Sequence: 1 agatcaactcacatcaaa.....gttttgaagataaagatg 719

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.2	8.4	9789	1 T41852	cDNA encoding Plas
2	56.4	7.8	5849	1 V33135	Plasmodium berghei
3	56	7.8	4590	1 N60472	Sequence encoding
4	56	7.8	19124	1 T72882	Plasmodium var-7 g
5	54.2	7.5	1186	1 V29580	Pythium oligandrum
6	53.2	7.4	53585	1 X20251	Borrelia burgdorfe
7	52.6	7.3	605	1 T31530	Human 3' apolipop
8	52.4	7.3	26811	1 X20253	Borrelia burgdorfe
9	52.2	7.3	1826	1 V74133	Orpinomyces cellul
10	52.2	7.3	1826	1 V29477	Orpinomyces cellul
11	52.2	7.3	2503	1 Q33480	PNPX30 xylanase cd
12	52	7.2	8920	1 Q52924	Carbamoyl-phosphat
13	51.8	7.2	605	1 T31530	Human 3' apolipop
14	51.8	7.2	3975	1 N81157	Malaria-specific g
15	51.8	7.2	3975	1 Q22999	SERP gene. Recombi
16	51.8	7.2	6124	1 Q33568	Sequence encoding
17	51.8	7.2	26811	1 X20253	Borrelia burgdorfe
18	51.4	7.1	9789	1 T41852	cDNA encoding Plas
19	51.4	7.1	19124	1 T72882	Plasmodium var-7 g
20	51.2	7.1	783	1 X20361	Borrelia burgdorfe
21	51.2	7.1	1711	1 V33136	Plasmodium berghei
22	51.2	7.1	1864	1 N71405	Sequence of ANS-1
23	51.2	7.1	3933	1 V29578	Pythium oligandrum
24	51.2	7.1	3933	1 V29578	Pythium oligandrum
25	51	7.1	1470	1 Q55185	MS-Lei610 Vector.
26	50.8	7.1	110000	1 V21209_11	Continuation (12 o
27	50.2	7.0	2418	1 Q27886	P.falciplarum GBP13
28	50.2	7.0	4970	1 Q49959	Impatiens Necrotic
29	50.2	7.0	6124	1 Q03568	Sequence encoding
30	50.2	7.0	8920	1 Q62924	Carbamoyl-phosphat
31	50	7.0	2104	1 Q25273	Sequence encoding
32	49.8	6.9	4467	1 T68648	PSRQ800 fragment 1
33	49.4	6.9	1611	1 Q12528	Thymidylate phosph

34	49.4	6.9	53585	1 X20251	Borrelia burgdorfe
35	49.2	6.8	1240	1 Q24177	Tox2a gene. DNA en
36	49.2	6.8	2503	1 O53480	PNPX30 xylanase cd
37	49	6.8	1671	1 Q24134	50 kD subunit of S
38	49	6.8	4098	1 T43682	Medium chain-speci
39	48.8	6.8	110000	1 V21209_12	P. falciplarum live
40	48.8	6.8	6152	1 T78867	Continuation (13 o
41	48.6	6.8	6152	1 T78867	P. falciplarum live
42	48.6	6.8	110000	1 V21209_13	Continuation (14 o
43	48.4	6.7	1982	1 N90225	Malaria-specific p
44	48.4	6.7	2763	1 N95079	Gmshp26-A heat sho
45	48.2	6.7	2849	1 X20293	Borrelia burgdorfe

## ALIGNMENTS

RESULT 1

T41852  
ID: T41852 standard; DNA: 9789 BP.  
AC T41852;  
DT 20-FEB-1997 (first entry)  
DE cDNA encoding Plasmodium falciparum erythrocyte membrane protein.  
KW Plasmodium falciparum; erythrocyte membrane protein; malaria;  
KW detection; identification; treatment; prevention; parasite; ss.  
OS Plasmodium falciparum MC type.  
FH Key Location/Qualifiers  
FT cds 326..9497  
FT /\*tag= a  
FT /\*product= Erythrocyte membrane protein  
FT misc\_feature 518..520  
FT /\*tag= b  
FT /\*transl\_except= GTA encodes Tyrosine  
FT misc\_feature 656..658  
FT /\*tag= c  
FT /\*transl\_except= ATT encodes Leucine  
FT misc\_feature 2909..2911  
FT /\*tag= d  
FT /\*transl\_except= AAC encodes Aspartic acid  
FT misc\_feature 3461..3463  
FT /\*tag= e  
FT /\*transl\_except= GAA encodes Glutamine  
FT misc\_feature 5546..5548  
FT /\*tag= f  
FT /\*transl\_except= CCT encodes Arginine  
FT misc\_feature 6254..6256  
FT /\*tag= g  
FT /\*transl\_except= AAT encodes Lysine  
FT misc\_feature 6257..6259  
FT /\*tag= h  
FT /\*transl\_except= ATA encodes Tyrosine  
FT misc\_feature 6263..6265  
FT /\*tag= i  
FT /\*transl\_except= AAC encodes Lysine  
FT misc\_feature 6269..6271  
FT /\*tag= j  
FT /\*transl\_except= TTC encodes Isoleucine  
FT misc\_feature 6272..6274  
FT /\*tag= k  
FT /\*transl\_except= ATA encodes Histidine  
FT misc\_feature 6275..6277  
FT /\*tag= l  
FT /\*transl\_except= ATT encodes Asparagine  
FT misc\_feature 6278..6280  
FT /\*tag= m  
FT /\*transl\_except= GGA encodes Tryptophan  
FT intron 7754..8478  
FT /\*tag= n

WO9633736-A1.

31-OCT-1996.

26-APR-1996; U05798.

27-APR-1995; US-430908.

PA (AFY-) AFFYMAX TECHNOLOGIES NV.

PI Baruch DI, Howard RJ, Pasloske BL;  
DR MPI; 96-497376/49.  
DR P-PSDB; W00384.

PT New Plasmodium falciparum erythrocyte membrane proteins - used to  
PT develop products for the diagnosis, treatment or prevention of  
PT malaria parasite infections

PS Disclosure: Figure 12; 149pp; English.

CC A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte  
CC membrane protein 1 (pfEMP1) or active fragments or analogues of that  
CC protein can be used in the treatment or prevention of symptoms of a  
CC malaria parasite infection. The polypeptides can inhibit, block or  
CC reverse the sequestration of erythrocytes in patients suffering from  
CC malaria. Nucleic acids derived from the pfEMP1 gene can be used as  
CC probes and primers to identify a Plasmodium falciparum parasite, the  
CC primers used to generate characteristic amplification patterns from  
CC different P. falciparum strains. Antibodies specifically  
CC immunoreactive with the pfEMP1 polypeptide or its fragments may be  
CC used in diagnosis of malaria infection. This sequence encodes the  
CC pfEMP1 protein of the MC type of Plasmodium falciparum. An  
CC alternative, truncated version of the coding sequence (a cDNA clone)  
CC is given in T41853.

CC Sequence 9789 BP: 4061 A; 1393 C; 1837 G; 2498 T;

SQ

Query Watch 8.4%; Score 60.2; DB 1: Length 9789;  
Best Local Similarity 48.2%; Pred. NO. 0.04;  
Matches 170; Conservative 0; Mismatches 183; Indels 0; Gaps

QY 53 aactatttgggttaaatattaatcattatttttaaagatatattaataagaataaaaga 112  
|| || || || || || || || || || || || || || || || || || || || ||  
Db 7900 AAAAAATTATTAAATAAATAAATAAANAANAANAATTCATAAAAAAAAAAAGAAATGCAAAA 7959

QY 113 tttttcaaaaaagtacaaaatatatttcattcatgatttttcacacatttgatttgtat 172  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 7960 TTTTATATAAAAAAATTTATTAATAAAATAATGATTATAAAAAAATTTATTAGAAA 8019

QY 173 aataaatatatttttttaattcttcaaaaaagtgtcgaaagaccttataggacatagtc 232  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 8020 TAAATATAAAAAATAATTTTATTAAATAAAAAANAAGAAAAAATAATGCTAAAAAATA 8079

QY 233 ttgttcgtttcacaaagcattccattcaatacatcaaaaaatatttaatactaag 292  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 8080 TATATATATATCATAAATATAAAAAANAATGTAAAGAANAATATATATTTATAAATAA 8139

QY 293 tagaatctcttctgtgagtggtgtaggaaccctggcattgaaacgagagagaag 352  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 8140 AAAAAAATTTAAAAAATAATGTTTTAAAAAATAATATATATATATAAATAAATAA 8199

QY 353 agtcgaacaccagagacaataaaaaagtatgcacaacaacaatacaaatcaaa 405  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 8200 TAAAAAATAATGTAANAANAANAANAATAAATAAATAAATAAATAAATAA 8252

RESULT 2

V33135/c

ID V33135 standard; DNA; 5849 BP.

AC V33135;

DT 07-OBC-1998 (first entry)

DE Plasmodium berghei-plastid PS1-PL470 gene.

KW Malaria; infection; therapy; diagnosis; vaccine; plastid;

KX PSI-PL470 gene; ds.

OS Plasmodium berghei ANKA strain.

PN WO9835057-A1.

PO 13-AUG-1998.

PF 05-FEB-1998; IB0212.

PR 26-SEP-1997; AU-009481.

PR 06-FEB-1997; AU-004953.

PR 21-APR-1997; AU-006329.

PA (MOLE-) INST MOLECULAR & CELL BIOLOGY.

PA (UYSI-) UNIV SINGAPORE NAT.

PI Kara AKU, Nelson JS, Tan TM, Tham JM, Ting RCY;

DR MPI; 98-447251/38.

PT Detecting Plasmodium infection from hybridisation with



CC parasite. This sequence can be used in the compositions of the invention.  
CC The compositions are for the treatment and prevention of malaria, and  
CC comprise either a nucleotide sequence or encoded polypeptide of the  
CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of  
CC genes having homology with conserved regions of DABP and SABB. The  
CC compositions are used for the treatment and prevention of malaria. They  
CC are also used in the preparation of vaccines for inducing a protective  
CC immune response in a mammal to Plasmodium merozoites (especially  
CC Plasmodium falciparum or Plasmodium vivax).  
SQ Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T;

Query Match 7.8%; Score 56; DB 1; Length 19124;  
Best Local Similarity 47.7%; Pred. No. 0.18;  
Matches 164; Conservative 0; Mismatches 180; Indels 0; Gaps 0

QY 54 attatttgggttaaaatcattcattttttaagatatattaataagaataaaaagt 113  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1264 ATTATAATGTAAGTAAATTATTAATAAAAATATATTGTATAACATACAAGACTTAAGAAGAAC 1323  
  
QY 114 tttttaaaaaatgataaaattatttcattcatgatgtttttccatcacattgatttgata 173  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1324 TATCAAACTCGTGATCTAATAGTATATATATATATATCTTTTTTTAAATTTGGTCTCT 1383  
  
QY 174 ataatatatttttttttttattcttcaaaaatggtgcgaagacacttattagacatagtct 233  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1384 CTTTTTTTTTTTTTAAAAATAATAAAATATTAATATATATTTTTTTTCATAATTTATATGAT 1443  
  
QY 234 tgtctgtttacaagaagcattcatcattttaatacatattaaaaaatattttaactaacagt 293  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1444 TTAGTATTTTAAATAATAAATAAATCTTTTAAAAAACCTTCAAACACATTTTGGCATAAAATA 1503  
  
QY 294 agaactctcttgtagtggtgtgggtagcgcaaccctgcatgaaacgagagaaagaga 353  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1504 ATATTAAATTAGTAGCACCATCTAGATAAAATTAGAGAGAAACGTAGAACATACCAGAAAAA 1563  
  
QY 354 gtcgaagaccagaagacaaataaaaaagtgatgcacaaacaaatca 397  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1564 ATTAGACAAAAGAAATATTACAAAAAATAATAAAAAATTAAATTA 1607

RESULT 5  
V29580/c ID V29580 standard; DNA; 1186 BP.

AC DT V29580;  
DE 04-SEP-1998 (first entry)  
DE Pythium oligandrum isolate 23-5 mitochondrial DNA partial sequence.  
KW Pythium oligandrum; phytopathogenic; mitochondrial DNA; fungus;  
KW soil saprophyte; pathogen; tomato; mycotoxic; plant protection;  
KW vegetable crop; ss.  
OS Pythium oligandrum.  
FH Key Location/Qualifiers  
FT misc\_feature 502..684  
ET /\*tag= a /note= "unique fragment claimed in claim 19"

FT PN W09816110-A1.  
PD PD 23-APR-1998.  
PF 10-OCT-1997; U18343  
PR 17-OCT-1996; US-731722.  
PA (UYFL) UNIV FLORIDA.  
PI Martin FN:  
PT WPI: 98-250977/22.  
PT Controlling phytopathogenic organisms with non-pathogenic Pythium  
PT isolate - for control of damping off caused by Pythium  
PS Claim 17; page 30; 4lpp; English.  
CC This is a partial nucleotide sequence of the mitochondrial DNA from  
CC a Pythium oligandrum isolate 23-5. Nucleotide sequences which are  
CC inverted repeats, flanked by PstI restriction sites from mitochondrial  
CC DNA from various P. oligandrum isolates are shown in V29576 to V29583.  
CC These Pythium isolate sequences are non-pathogenic and can be used in a  
CC method for controlling phytopathogenic organisms where the organisms are  
CC contacted with such a Pythium isolate. The Pythium isolates are used to  
CC control fungi, specifically pathogenic Pythium species (but possibly also

CC other soil-borne pathogens), particularly for protecting plants  
CC (seedlings, transplants or vegetable crops such as tomato) against  
CC damping off, especially after transplanting into open fields. The  
CC isolate sequences are useful as sources of probes for identification of  
CC particular isolates. The isolates are widely distributed in nature, with  
CC a similar ecology to pathogenic species. They produce large quantities of  
CC oospore inoculum on liquid or solid substrates and are tolerant of  
CC several commonly used fungicides. A single application at the greenhouse  
CC stage will protect plants after transplanting.  
SQ Sequence 1186 BP; 469 A; 121 C; 125 G; 471 T;

Query Match 7.5%; Score 54.2; DB 1; Length 1186;  
Best Local Similarity 59.3%; Pred. NO. 0.38;  
Matches 112; Conservative 0; Mismatches 73; Indels 4; Gaps 1;

QY 49 tactaatttttgggttaataatcattatttttaagataatttaataagaaattaa 108  
DB 595 TATTGAAGAAATTTCTTATACTACTTAAAGATTATAGGATCTGTGTAAATCCAG 536  
QY 109 aegatttttttaaaaaatgtataaaattattatttcattgatttttcacacattgatt 168  
DB 535 ATCATTATTATAAATAATAAAGTTT---TTCAATATTATAATAAATAATATTC 480  
QY 169 tgataataatatttttttaattcttcaaaaaatgtgcgaagcactattagacat 228  
DB 479 TTATTAGAGTATTTCATTAAATTTTATTTTAAAGATTATATATCTTTAAAAAGATAT 420  
QY 229 agcttggtt 237  
DB 419 AAATTAATT 411

## RESULT 6

ID X20251 standard; DNA; 53585 BP.  
AC X20251;  
DT 04-MAY-1999 (first entry)  
DE Borrelia burgdorferi polynucleotide sequence #4.  
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;  
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
KW infection; diagnosis; characterisation; detection; ds.  
OS Borrelia burgdorferi.  
PN W09858943-AL.  
PD 30-DEC-1998.  
PF 18-JUN-1998; U12764.  
PR 03-SEP-1997; US-057483.  
PR 20-JUN-1997; US-050359.  
PR 22-JUL-1997; US-053344.  
PR 22-JUL-1997; US-053377.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (MED-) MEDIMUNE INC.  
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,  
PI White OR;  
DR WPI; 99-081217/07.  
PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
PT products for the detection, diagnosis, characterisation, prevention  
PT and therapy of infections, particularly Lyme disease  
PS Claim 1; Page 801-831; 1128pp; English.  
CC X20248 to X20402 represent polynucleotide sequences isolated from  
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for  
CC the detection, diagnosis, characterisation, prevention and therapy of  
CC Bb infections, e.g. Lyme disease. They can also be used for the  
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs  
CC to a family of mobile, spiral-shaped bacteria called Spirochetes.  
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
CC Lyme disease.  
SQ Sequence 53585 BP; 19235 A; 8081 C; 7001 G; 19266 T;

Query Match 7.4%; Score 53.2; DB 1; Length 53585;  
Best Local Similarity 52.2%; Pred. NO. 0.46;

Matches 118; Conservative 0; Mismatches 108; Indels 0; Gaps 0;  
QY 18 aacataacatgatactctcttaccacatactactatttttgggttaaaatattaaac 77  
DB 33384 AATATTAAATAATTAATAATAATTTAAACAAATTAATTTATATTAATAAATG 33325  
QY 78 attatttttaagataataattgaagaatttaaaagatttttaaaaaatgtacaaatta 137  
DB 33324 CAAATTTTGTAAAAAATAAAAAATAATTAATCAATCTTTTAAAGATTTTGAATAATTT 33265  
QY 138 tattattcatgatttttcacacattgattgataataataatttttttaattct 197  
DB 33264 TTTTAAAGTTTATTTTGGAAATAATTAATTCATGATGATTATATATTTTATTA 33205  
QY 198 taataaatgtgcgaagcactatttagacatagcttctgtctgtt 243  
DB 33204 TTTTACCACCTAAGGACTCTATTATGAAAAACAGATTTTCTCTATAT 33159

## RESULT 7

ID T31530 standard; cDNA; 605 BP.  
AC T31530;  
DT 15-SEP-1996 (first entry)  
DE Human 3' apolipoprotein B SAR element clone Rh32.  
KW Erythropoietin; EPO; anaemia; gene therapy; vector;  
KW scaffold attachment region; SAR element; apolipoprotein B;  
KW transgenic animal; ss.  
OS Homo sapiens.  
PN W09619573-AL.  
PD 27-JUN-1996.  
PF 18-DEC-1995; CA0696.  
PR 19-DEC-1994; US-358918.  
PA (CANG-) CANGENE CORP.  
PI Delcuve G;  
DR WPI; 96-309587/31.  
PT Recombinant DNA molecule expressing mammalian erythropoietin  
PT useful to transform cell lines, and for gene therapy, e.g. of  
PT anaemia and other red blood cell disorders  
PS Claim 7; Page 59-60; 84pp; English.  
CC Human apolipoprotein B (apoB) scaffold attachment region (SAR)  
CC element clones Rh32 (T31530) and Rh10 (T31531) respectively carry  
CC the 3' human apoB SAR element and the distal 1212 bp 5' human apoB  
CC SAR element and 1317 bp proximal sequence. These SAR elements  
CC co-map with the boundaries of the human apoB gene chromatin  
CC domain. A novel recombinant DNA molecule adapted for transfection  
CC of a host cell comprises an erythropoietin (EPO) cDNA (T31529) or  
CC genomic clone (T31532) operably linked to an expression control  
CC sequence and to the 5' and 3' SAR elements. The SAR elements  
CC increase expression of the recombinant EPO in stable, long-term  
CC mammalian cell cultures.  
SQ Sequence 605 BP; 278 A; 18 C; 35 G; 274 T;

Query Match 7.3%; Score 52.6; DB 1; Length 605;  
Best Local Similarity 55.1%; Pred. NO. 0.7;  
Matches 124; Conservative 0; Mismatches 99; Indels 2; Gaps 1;

QY 70 tattaatcattttttaagataataaagaatttaaaagatttttaaaaaatgta 129  
DB 25 TATAATTAATAATTTATATAATTAATAATTTATTAATAATAATTTATTAATAATTT 84  
QY 130 taaaattattattcatgatttttcacacattgatttttgataataataatttttt 189  
DB 85 TATAATTAATAATTTATATAATTAATAATAATAATAATAATAATAATAATAAT 142  
QY 190 taattctttaaaaaatgtgcgaagcactatttagacatagcttctgtctgttcaaaaa 249  
DB 143 TTTTATAATAATAATTTATATAATTAATAATAATAATAATAATAATAATAAT 202  
QY 250 gcattcatcatttaacacatttaaaaaatattttaactaacagta 294  
DB 203 TTTTATAATAATAATAATTTATATAATAATAATAATAATAATAATAATAATA 247

```

RESULT      8
X20253
ID X20253 standard; DNA; 26811 BP.
AC X20253;
DT 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #6.
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
PN W09858943-A1.
PD 30-DEC-1998.
PF 18-JUN-1998: U12764.
PR 03-SEP-1997: US-057483.
PR 20-JUN-1997: US-050359.
PR 22-JUL-1997: US-053344.
PR 22-JUL-1997: US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI White OR;
DR WPI: 99-081217/07.
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1; Page 851-867; 1129pp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (BB). Products derived from BB can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC BB infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
SQ Sequence 26811 BP; 8978 A; 3456 G; 3596 G; 10780 T;

Query Match 7.3%; Score 52.4; DB 1; Length 26811;
Best Local Similarity 52.3%; Pred. No. 0.63;
Matches 116; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 76 tcattatttttaagatatttaataagaataaaagatttttttaaaaaaatgtataaat 135
DB 14768 TAATTATTATTCTATTGTTAAATAGATTATTATTAGATTATTATTATTATTATTAATAAAAAA 14827

QY 136 tatattattcatgattttcatcacattgttattgataataataatatttttttaattt 195
DB 14828 TAATTATTATTGAAATTTTAAAAAATAGAAATTAATTTCTAACTAAATTTTCATATTAA 14887

QY 196 cttaaaaaaatgtgcgaagacattattagacatagctgtgtgttttcaaaaagcattc 255
DB 14888 CACTTCCTTGTGGCAATAAAAAGTTCTATACAGAAGCTTCTTCGTTTAAAAACACCTTT 14947

QY 256 atcatttaatacattaaaaaatttttaataactaacagtagaa 297
DB 14948 TAAAGAGTTTAAAAAAGCTTTCTTATATATATTCTTCAACAAAAAGCA 14989

RESULT      9
V37413
ID V37413 standard; cDNA to mRNA; 1826 BP.
AC V37413;
DT 14-SEP-1998 (first entry)
DE Orpinomyces cellulase CelB cDNA.
KW Cellulase; endoglucanase; CelB gene; ds.
OS Orpinomyces sp. strain PC-2.
FH Key Location/Qualifiers
FT CDS
FT 69..1484
FT /*tag= a
FT /*note= "Claim 5"
PT New recombinant DNA encoding Orpinomyces cellulase protein - useful

```

```

PN W09814597-A1.
PD 09-APR-1998.
PF 03-OCT-1997: U18008.
PR 04-OCT-1996; US-027883.
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
PI Chen H, Li X, Ljungdahl LG;
DR WPI: 98-240096/21.
DR P-PSDB: W56742.
PT New recombinant DNA encoding Orpinomyces cellulase protein - useful
PT for, e.g. producing recombinant Orpinomyces cellulase in host cell
PS Claim 5; Page 38-40; 69pp; English.
CC This cDNA clone includes a claimed coding region for cellulase celB
CC (see W56742) of the anaerobic bovine rumen fungus Orpinomyces sp.
CC PC-2. It was obtained by screening a PC-2 cDNA library for clones
CC active on remazol brilliant blue-carboxymethylcellulose. The
CC encoded cellulase has endoglucanase, but not cellobiohydrolase,
CC activity. CelA and celC genes (see V29472-73), also obtained from
CC Orpinomyces sp. PC-2, encode cellulases having both activities
CC (see W56738-39). Recombinant DNA molecules encoding Orpinomyces
CC cellulase proteins are claimed, as well as recombinant cells
CC selected from Saccharomyces cerevisiae, Escherichia coli,
CC Aspergillus, Trichoderma reesei, Pichia, Penicillium, Streptomyces
CC or Bacillus, and a method for producing recombinant cellulase by
CC culturing these host cells.
SQ Sequence 1826 BP; 654 A; 236 C; 313 G; 623 T;

Query Match 7.3%; Score 52.2; DB 1; Length 1826;
Best Local Similarity 49.8%; Pred. No. 0.76;
Matches 132; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 10 tcacatccaaacaataacatggatctcttaccatcatactaatattttgggttaa 69
DB 1501 TTAACAACATAAATAATATTATTAGTAAATAAAGAAATAAATTTTAAAAAATAT 1560

QY 70 tattaatcattttttaagatatttaataagaataaaagatttttttaaaaaaatgta 129
DB 1561 ATTATATTATTGTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATA 1620

QY 130 taaaattattattattcatgattttttcatcacattgttattgataataataatttttt 189
DB 1621 AAATATATACATAACAAAAAGTAAAAAATTTTAAATTTTAGTATTGTATATAATTTT 1680

QY 190 taatttttaaaaaatgttcgaagacattattagacatagctgttctgtttacaaaa 249
DB 1681 AAAAAGTTTAATAATAAGTAAAAAATAAATAATAATAATAATAATAATAATAATAATA 1740

QY 250 gcattcatcatatttaacatacaataaa 274
DB 1741 AGAAATAATAATAAATACITTTAAA 1765

RESULT 10
V29477
ID V29477 standard; cDNA to mRNA; 1826 BP.
AC V29477;
DT 21-SEP-1998 (first entry)
DE Orpinomyces cellulase CelB cDNA.
KW Cellulase; endoglucanase; CelB gene; ds.
OS Orpinomyces sp. strain PC-2.
FH Key Location/Qualifiers
FT CDS
FT 69..1484
FT /*tag= a
FT /*note= "Claim 5"
PT New recombinant DNA encoding Orpinomyces cellulase protein - useful

```

PT for, e.g. producing recombinant *Orpinomyces* cellulase in host cell  
 PS Claim 5; Page 38-40; 69pp; English.  
 CC This cDNA clone includes a claimed coding region for cellulase celB  
 CC (see W56742) of the anaerobic bovine rumen fungus *Orpinomyces* sp.  
 CC PC-2. It was obtained by screening a PC-2 cDNA library for clones  
 CC active on remazol brilliant blue-carboxymethylcellulose. The  
 CC encoded cellulase has endoglucanase, but not cellobiohydrolase,  
 CC activity. CelA and celC genes (see V29472-73), also obtained from  
 CC *Orpinomyces* sp. PC-2, encode cellulases having both activities  
 CC (see W56738-39). Recombinant DNA molecules encoding *Orpinomyces*  
 CC cellulase proteins are claimed, as well as recombinant cells  
 CC selected from *Saccharomyces cerevisiae*, *Escherichia coli*,  
 CC *Aspergillus*, *Trichoderma reesei*, *Pichia*, *Penicillium*, *Streptomyces*  
 CC or *Bacillus*, and a method for producing recombinant cellulase by  
 CC culturing these host cells.  
 CC Sequence 1826 BP; 654 A; 236 C; 313 G; 623 T;  
 SQ

Query Match 7.3%; Score 52.2; DB 1; Length 1826;  
 Best Local Similarity 49.8%; Pred. No. 0.76;  
 Matches 132; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 10 tccatcccaaacataacatgagatctctctaccatcactactactattttgggttaa 69  
 DB 1501 TTACAACATAATAAATTTAGTAAATAAAGAAAGAAATAATTTTAAATAATAT 1560  
 QY 70 tattaatcatttttaagatttaataagaatttaaaagatttttttaaaataatgta 129  
 DB 1561 ATTATATATATGTTATTAATAATAATAATAATAATAATAATAATAATAATAATA 1620  
 QY 130 taaatattattattcatgattttttcaccattgattttgataataataattttttt 189  
 DB 1621 AAATATATACATAACAAAGTAAATAATTAATAATTTTAAATTTTAAATTTTATT 1680  
 QY 190 taattttttaaataatgttcagacactcattagacatgcttctgttttcaaaaaa 249  
 DB 1681 AAAAAGTTTAATAAGTAAATAAATAAATAATAATAATAATAATAATAATAATAA 1740  
 QY 250 gcattcatcatttaacacataaaaa 274  
 DB 1741 AGAATAATAATAATACTTTTAAAA 1765

RESULT 11  
 Q53480  
 ID Q53480 standard; cDNA; 2503 BP.  
 AC Q53480;  
 DT 30-JUN-1994 (first entry)  
 DE pNPX30 xylanase cDNA.  
 KW xylanase; ruminant animals; fungus; pulp; bagasse;  
 KW feedstock; rumen; plant fibre; ss.  
 OS Neocallimastix patriciarum.  
 FH Key Location/Qualifiers  
 FT cds 1.1935  
 FT /\*tag= a  
 FT /product= xylanase.  
 PN W09325671-A.  
 PD 23-DEC-1993.  
 PF 17-JUN-1993; AU0294.  
 PR 17-JUN-1992; AU-002985.  
 PR 29-JUN-1992; AU-003238.  
 PR 01-APR-1993; AU-008100.  
 PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.  
 PI Xue GJ;  
 DR WPI; 94-007529/01.  
 DR P-PSDB; R44529.  
 PT New recombinant fungal xylanase - used for hydrolysis of xylan in  
 PT food and pulp and paper industries and for improving ruminant  
 PT feed efficiency  
 PS Claim 9; Figure 3; 45pp; English.  
 CC The cloned xylanase coding sequence is derived from an anaerobic  
 CC rumen fungus. The xylanase has high specific activity for the  
 CC hydrolysis of xylan. It can be used for treating pulps in the pulp

CC and paper industry, for treating bagasse for more efficient disposal  
 CC or for the treatment of feedstock to improve nutritional value.  
 CC Genetically modified xylanase genes can also be used for the  
 CC modification of rumen bacteria to improve plant fibre utilisation by  
 CC ruminants. 2503 BP; 893 A; 389 C; 517 G; 704 T;  
 SQ

Query Match 7.3%; Score 52.2; DB 1; Length 2503;  
 Best Local Similarity 51.6%; Pred. No. 0.75;  
 Matches 144; Conservative 0; Mismatches 133; Indels 2; Gaps 1;

QY 18 aaacataacatgatatctctctaccatcactactactattttgggttaaatatttc 77  
 DB 2070 AAATTTTAAATGTAAAAATTTAAAAATAACAATTTGTAAAAAAATGAAGCAATTATG 2129  
 QY 78 attatttttaagatatattaagaataataaagatttttttaaaaaatgtataaatta 137  
 DB 2130 AAAAATTTAAATGTAAAAATTTAAAAATAACAATTTGTAAAAAAATTAAGAAATTATAA 2189  
 QY 138 tattattcatgatttttcacattgatttttgataa--taaatatatttttttaattt 195  
 DB 2190 AAAAATAAAGAATTTATGAATAATTTAAATGTAAAAATTTAAAAATATATATTTTAAAT 2249  
 QY 196 cttaaaaaatgttcgaagacacttattagacatgcttctgttttccaaagaatttc 255  
 DB 2250 AATAAAGAATTTATGAATAATTTAAATATAAATAATAAAGTGTGTTTATAGTAAAAATA 2309  
 QY 256 atcatttaacatttaaaaaataatttaataactaacagta 294  
 DB 2310 AAAAATTTATGAATAATTTTAAATATAAATAATAAATAATAAATAATAAATA 2348

RESULT 12  
 Q62924/C  
 ID Q62924 standard; cDNA; 8920 BP.  
 AC Q62924;  
 DT 06-DEC-1994 (first entry)  
 DE Carbamoyl-phosphate-synthetase II.  
 KW Carbamoyl-phosphate-synthetase II; CPSII; pscCPSII gene;  
 KW malaria; ss.  
 OS Plasmodium falciparum.  
 FH Key Location/Qualifiers  
 FT cds 1226.8401  
 FT /\*tag= a  
 FT /EC\_number= 6.3.5.5  
 PN W09412643-A.  
 PD 09-JUN-1994.  
 PF 02-DEC-1993; AU0617.  
 PR 03-DEC-1992; AU-006206.  
 PR 16-DEC-1992; AU-006380.  
 PA (UNIX ) UNISEARCH LTD.  
 PI Flores WJ, Osullivan WJ, Stewart TS;  
 DR WPI; 94-200271/24.  
 DR P-PSDB; R55694.  
 PT Nucleic acid encoding carbamoyl phosphate synthetase II -  
 PT isolated from plasmodium falciparum, used to develop prods. for  
 PT the treatment of malaria.  
 PS Disclosure; Page 6-16; 31pp; English.  
 CC The cDNA sequence encoding the carbamoyl-phosphate-transferase II  
 CC (CPSII) of plasmodium falciparum was determined. The cDNA encodes  
 CC a protein that includes 2 insert sequences not found in other CPSII  
 CC proteins. The first separates the putative structural subdomain and  
 CC the glutaminase subdomain of the glutamine-amidotransferase subunit  
 CC of CPSII, while the second separates 2 ATP binding subdomains of the  
 CC CPSII subunit, CPsA and CPsB.  
 SQ Sequence 8920 BP; 8836 A; 774 C; 1232 G; 3078 T;

Query Match 7.2%; Score 52; DB 1; Length 8920;  
 Best Local Similarity 49.8%; Pred. No. 0.76;  
 Matches 133; Conservative 0; Mismatches 135; Indels 0; Gaps 0;





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: September 2, 2000, 02:55:46 : Search time 5217.98 Seconds  
(without alignments)  
245.918 Million cell updates/sec

Title: US-09-464-528-16  
Perfect score: 719

Sequence: 1 agatcaaacacacatccaaa.....gttttgaagtataaagatg 719

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba1.\*

2: gb\_ba2.\*

3: gb\_cm.\*

4: gb\_ov.\*

5: gb\_pat.\*

6: gb\_ph.\*

7: gb\_pl1.\*

8: gb\_pl2.\*

9: gb\_pr1.\*

10: gb\_pr2.\*

11: gb\_pr3.\*

12: gb\_ro.\*

13: gb\_sts.\*

14: gb\_sy.\*

15: gb\_un.\*

16: em\_fun.\*

17: em\_hum1.\*

18: em\_hum2.\*

19: em\_in.\*

20: em\_cm.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_sy.\*

29: em\_un.\*

30: em\_v1.\*

31: gb\_htg1.\*

32: gb\_htg2.\*

33: gb\_in1.\*

34: gb\_in2.\*

35: em\_ba1.\*

36: em\_ba2.\*

37: em\_hum3.\*

38: em\_hum4.\*

39: gb\_pr4.\*

40: gb\_htg3.\*

41: gb\_htg4.\*

42: gb\_htg5.\*

43: gb\_htg6.\*

44: gb\_htg7.\*  
45: em\_htg1.\*  
46: em\_htg2.\*  
47: em\_htg3.\*  
48: em\_hum5.\*  
49: gb\_pl3.\*  
50: gb\_br5.\*  
51: gb\_htg8.\*  
52: gb\_htg9.\*  
53: gb\_htg10.\*  
54: gb\_htg11.\*  
55: gb\_htg12.\*  
56: gb\_htg13.\*  
57: gb\_htg14.\*  
58: gb\_in3.\*  
59: gb\_htg15.\*  
60: gb\_htg16.\*  
61: gb\_htg17.\*  
62: em\_htg4.\*  
63: em\_htg5.\*  
64: em\_htg6.\*  
65: em\_htg7.\*  
66: em\_hum6.\*  
67: gb\_htg18.\*  
68: gb\_htg19.\*  
69: gb\_htg20.\*  
70: gb\_htg21.\*  
71: gb\_htg22.\*  
72: gb\_htg23.\*  
73: gb\_htg24.\*  
74: gb\_htg25.\*  
75: gb\_htg26.\*  
76: gb\_htg27.\*  
77: gb\_htg28.\*  
78: gb\_htg29.\*  
79: gb\_htg30.\*  
80: gb\_htg31.\*  
81: gb\_v11.\*  
82: gb\_v12.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74	10.3	144759	10 HS352A20	AL021939 Homo sapi
2	70.2	9.8	153267	57 AC011212	AC011212 Homo sapi
3	70	9.7	104992	41 AC005504	AC005504 Plasmodiu
4	70	9.7	130281	60 AC004157	AC004157 Plasmodiu
5	69.8	9.7	349919	54 AC008576	AC008576 Homo sapi
6	69.6	9.7	164119	75 AC026640	AC026640 Homo sapi
7	69.2	9.6	179510	71 AC013820	AC013820 Homo sapi
8	69	9.6	48532	51 AC023371	AC023371 Homo sapi
9	68.4	9.5	975	33 DD087514	U87514 Dictyosteli
10	68.2	9.5	80920	60 AC006278	AC006278 Plasmodiu
11	68.2	9.5	121024	8 CHMPXX	X04465 Marchantia
12	67.8	9.4	910	13 CNS01G8P	AL142826 Anopheles
13	66.8	9.3	99263	32 CNS01DX9	AL139177 Homo sapi
14	66.8	9.3	152209	11 HS1108D11	AL034419 Human DNA
15	66.6	9.3	1867	7 MTSCAJ23	AJ223323 Saccharom
16	66.4	9.2	149752	11 AC004616	AC004616 Homo sapi
17	66.4	9.2	170427	41 AC006095	AC006095 Homo sapi
18	66.2	9.2	167271	53 AC006269	AC006269 Homo sapi
19	66.2	9.2	226345	39 AC005406	AC005406 Homo sapi
20	66	9.2	158131	68 AC009277	AC009277 Homo sapi
21	65.8	9.2	2982	33 DDISGSPA	M3862 Dictyosteli
22	65.8	9.2	67970	33 PFMALIP3	AL031746 Plasmodiu
23	65.4	9.1	690	7 MISC10	V00691 Two yeast m
24	65.4	9.1	716	8 YSCMTTGT1	J01533 Yeast (S.ce

25 65.2 9.1 68879 72 AC031982 Homo sapi  
 26 65.2 9.1 287731 59 AC008620 Homo sapi  
 27 65 9.0 2305 33 DMTRNA X54011 D. teissleri  
 28 65 9.0 183638 77 AC009653 Homo sapi  
 29 65 9.0 216406 52 AC018919 Homo sapi  
 30 64.8 9.0 161891 55 AC008206 Drosophila  
 31 64.8 9.0 163878 52 AC010178 Homo sapi  
 32 64.8 9.0 173893 57 AC021553 Homo sapi  
 33 64.4 9.0 217242 73 AC016222 Homo sapi  
 34 64.2 8.9 153267 57 AC011212 Homo sapi  
 35 64.2 8.9 174434 42 AC011115 Homo sapi  
 36 64 8.9 1398 8 SCU46121 U46121 Saccharomyc  
 37 64 8.9 2806 7 YSCMTG07 L38891 Saccharomyc  
 38 64 8.9 14867 34 AE001398 AE001398 Plasmid  
 39 64 8.9 85779 8 SCE011856 AJ011856 Saccharom  
 40 64 8.9 106650 39 AC007708 Homo sapi  
 41 63.8 8.9 6591 7 YSCMTG06 L36890 Saccharomyc  
 42 63.8 8.9 181650 79 AC009760 Homo sapi  
 43 63.8 8.9 193372 53 AC022266 Homo sapi  
 44 63.8 8.9 204951 60 AC005505 Plasmid  
 45 63.6 8.8 145932 59 AC024387 Homo sapi

## ALIGNMENTS

## RESULT

## HS352A20

## LOCUS

## DEFINITION

HS352A20 144759 bp DNA PRI 23-NOV-1999  
 Homo sapiens DNA sequence from PAC 352A20 on chromosome  
 6q24.1-25.1. Contains a pseudogene similar to yeast, bacterial,  
 worm and slime mold hypothetical genes, and a gene coding for an  
 aldehyde dehydrogenase family protein. Contains ESTs, STSs and  
 GSSs, complete sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Submitted (07-MAY-1998) sanger.ac.uk/HGP/Chr6/) Sanger Centre,  
 Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On May 18, 1998 this sequence version replaced gi:2909620.  
 IMPORTANT: This sequence is the entire insert of clone 352A20.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variations annotated may not be found in the sequence submission  
 corresponding to the overlapping clone as we submit sequences with  
 only a small overlap as described above.  
 This sequence was generated from part of bacterial clone contigs of  
 human chromosome 6, constructed by the Sanger Centre chromosome 6  
 mapping group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr6/  
 This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.  
 The true left end of clone 352A20 is at 1 in this sequence. The  
 true right end of clone 352A20 is at 144759.  
 352A20 is from the library RPC13 constructed at the Roswell Park  
 Cancer Institute by the group of Pieter de Jong.  
 For further details see http://bacpac.med.buffalo.edu/  
 Location/Qualifiers  
 1..144759  
 /organism="Homo sapiens"

/db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="q24.1-25.1"  
 /clone="RP3-352A20"  
 /clone\_lib="RPC1-3"  
 3..211  
 /note="AluSc repeat: matches 77..285 of consensus;  
 incomplete repeat"  
 334..392  
 /note="MER45 repeat: matches 4..62 of consensus"  
 662..959  
 /note="AluSc repeat: matches 1..299 of consensus"  
 1362..1544  
 /note="L1PB3 repeat: matches 640..811 of consensus"  
 1546..1836  
 /note="AluSg repeat: matches 299..1 of consensus"  
 1839..1937  
 /note="L1MA4A repeat: matches 941..1046 of consensus"  
 2448..2745  
 /note="AluJb repeat: matches 3..301 of consensus"  
 3863..4153  
 /note="AluSg repeat: matches 1..293 of consensus"  
 4897..4960  
 /note="MIR2 repeat: matches 80..146 of consensus"  
 7345..7647  
 /note="AluSg repeat: matches 303..1 of consensus"  
 8224..8333  
 /note="MIR repeat: matches 178..66 of consensus"  
 8763..8945  
 /note="AluY repeat: matches 300..104 of consensus;  
 incomplete repeat"  
 8955..9012  
 /note="AluYb8 repeat: matches 27..83 of consensus;  
 incomplete repeat"  
 9495..9794  
 /note="AluJb repeat: matches 1..297 of consensus"  
 9797..10088  
 /note="AluX repeat: matches 1..301 of consensus"  
 11019..11210  
 /note="MIR repeat: matches 7..199 of consensus"  
 11212..11507  
 /note="AluX repeat: matches 1..296 of consensus"  
 13312..13612  
 /note="AluY repeat: matches 301..1 of consensus"  
 14187..14497  
 /note="AluY repeat: matches 1..301 of consensus"  
 14952..15077  
 /note="MIR repeat: matches 217..84 of consensus"  
 15302..15475  
 /note="MER5B repeat: matches 1..178 of consensus"  
 15829..16294  
 /note="MLTIC repeat: matches 1..466 of consensus"  
 17194..17481  
 /note="AluJo repeat: matches 1..302 of consensus"  
 17715..18093  
 /note="MSTB repeat: matches 4..361 of consensus"  
 18095..18395  
 /note="AluY repeat: matches 301..1 of consensus"  
 18397..18461  
 /note="MSTD repeat: matches 330..394 of consensus"  
 20030..20334  
 /note="AluSc repeat: matches 1..299 of consensus"  
 23448..23750  
 /note="AluX repeat: matches 1..300 of consensus"  
 24157..24456  
 /note="AluSg repeat: matches 1..300 of consensus"  
 29337..29632  
 /note="AluSg repeat: matches 296..1 of consensus"  
 30246..30393  
 /note="LOR1 repeat: matches 1..149 of consensus"  
 30394..30691  
 /note="AluX repeat: matches 1..299 of consensus"  
 30692..30789







Db 12406 ATATAATTATAATATAA 12388

RESULT	4
AC0004157/c	DNA
LOCUS	130281 bp
DEFINITION	Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.
ACCESSION	AC0004157
VERSION	AC0004157.6 GI:7243830
KEYWORDS	HTG; HTGS-PHASE1.
SOURCE	malaria parasite P. falciparum,
ORGANISM	Plasmodium falciparum
REFERENCE	Eukaryota: Alveolata: Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	1 (bases 1 to 130281) Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Tamaki,T., Kurdi,O.B., Conway,A.B. and Davis,R.W.
TITLE	Plasmodium falciparum 3D7 chromosome 12
JOURNAL	Unpublished
REFERENCE	2 (bases 1' to 130281)
AUTHORS	Hyman,R.W., Qin,F.; Fung,E.L., Conway,A.B. and Davis,R.W.
TITLE	Direct Submission
JOURNAL	Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304. USA

```

COMMENT
On Mar 15, 2000 this sequence version replaced gi:6652498.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1. 67262: contig of 67262 bp in length
* 67263 67462: gap of unknown length
* 67463 82485: contig of 13023 bp in length
* 82486 82685: gap of unknown length
* 82686 130281: contig of 47596 bp in length.
* Location/Qualifiers
* 1. 130281
* /organism="Plasmodium falciparum"
* /db_xref="taxon:5833"
* /chromosome="12"
FEATURES
source

```

BASE COUNT	52250 a	11780 c	11855 g	53996 t	400 others
Query Match	9.74;	Score 70;	DB 60;	Length 130281;	
Best Local Similarity	57.04;	Pred. No. 0.36;			
Matches 147;	Conservative	0;	Mismatches 110;	Indels 1;	Gaps
Qy	50	actaattattttgggttaaatattataatcattatttttaagatatttaattaaagaaataaa-	108		
Db	99223	ATTAATTAATATATATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT			99164
Qy	109	aagattttttaaataaagtataaattatattatctatctatgatttttcacacatttgatt	168		
Db	99163	TTAAATTTAAATAAATAATATATATATATAATAATAATAATAATAATAATAATAATAAT			99104
Qy	169	tgataataaaatatttttttaatttcttaaaaaagtgtgcaagacacatttagacat	228		
Db	99103	TAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT			99044
Qy	229	agtcgttctcgtttcacaaaagcatccatcatttaatacatataaaaaattattaatacta	288		
Db	99043	ATTTTATTTTATTTTAAATTAATTAATTAATAATAATAATAATAATAATAATAATAAT			98984
Qy	289	acagtagaatactctctgt	306		

Db 98983 AATAAATAATTATTAT 98966

## RESULT 5

AC008576/c

## LOCUS

DEFINITION

AC008576 349919 bp DNA HTG 18-FEB-2000  
Homo sapiens chromosome 19 clone CTC-557J18, WORKING DRAFT  
SEQUENCE, 56 unordered pieces.

## ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AC008576 349919 bp DNA HTG 18-FEB-2000  
Homo sapiens chromosome 19 clone CTC-557J18, WORKING DRAFT  
SEQUENCE, 56 unordered pieces.

ACCESSION  
VERSION AC008576.3 GI:6997021  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
1 (bases 1 to 349919) Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE DOE Joint Genome Institute.  
JOURNAL Sequencing of Human Chromosome 19  
REFERENCE Unpublished  
2 (bases 1 to 349919)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Feb 18, 2000 this sequence version replaced gi:6601030.

-----Genome Center  
Center: Joint Genome Institute.  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>

-----Summary Statistics  
Consensus quality: 285737 bases at least Q40  
Consensus quality: 300872 bases at least Q30  
Consensus quality: 311585 bases at least Q20  
Estimated insert size: 349919; sum-of-contigs estimation  
Estimated insert size: 314630; agarose-fp estimation  
Quality coverage: 1.90x in Q20 bases; agarose-fp estimation  
Quality coverage: 1.71x in Q20 bases; sum-of-contigs estimation

-----  
NOTE: This is a 'working draft' sequence. It currently  
consists of 56 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 3011: contig of 3011 bp in length  
gap of unknown length  
3012 6583: contig of 3572 bp in length  
gap of unknown length  
6584 9679: contig of 3096 bp in length  
gap of unknown length  
9680 11837: contig of 2158 bp in length  
gap of unknown length  
11838 13926: contig of 2089 bp in length  
gap of unknown length  
13927 16002: contig of 2076 bp in length  
gap of unknown length  
16003 18305: contig of 2303 bp in length  
gap of unknown length  
18306 20541: contig of 2236 bp in length  
gap of unknown length  
20542 23415: contig of 2874 bp in length  
gap of unknown length  
23416 25390: contig of 1975 bp in length  
gap of unknown length  
25391 27927: contig of 2537 bp in length  
gap of unknown length  
27928 30114: contig of 2187 bp in length  
gap of unknown length  
30115 32716: contig of 2602 bp in length  
gap of unknown length  
32717 34880: contig of 2164 bp in length  
gap of unknown length

34881 36992: contig of 2112 bp in length  
gap of unknown length  
36993 40002: contig of 3010 bp in length  
gap of unknown length  
40003 42499: contig of 2497 bp in length  
gap of unknown length  
42500 44537: contig of 2038 bp in length  
gap of unknown length  
44538 46636: contig of 2099 bp in length  
gap of unknown length  
46637 49202: contig of 2566 bp in length  
gap of unknown length  
49203 51194: contig of 1992 bp in length  
gap of unknown length  
51195 54232: contig of 3038 bp in length  
gap of unknown length  
54233 56311: contig of 2079 bp in length  
gap of unknown length  
56312 59177: contig of 2866 bp in length  
gap of unknown length  
59178 61728: contig of 2551 bp in length  
gap of unknown length  
61729 65138: contig of 3410 bp in length  
gap of unknown length  
65139 68509: contig of 3371 bp in length  
gap of unknown length  
68510 72140: contig of 3631 bp in length  
gap of unknown length  
72141 74177: contig of 2037 bp in length  
gap of unknown length  
74178 76657: contig of 2480 bp in length  
gap of unknown length  
76658 81120: contig of 4463 bp in length  
gap of unknown length  
81121 83944: contig of 2824 bp in length  
gap of unknown length  
83945 88969: contig of 5025 bp in length  
gap of unknown length  
88970 94016: contig of 5047 bp in length  
gap of unknown length  
94017 99820: contig of 5804 bp in length  
gap of unknown length  
99821 103739: contig of 3919 bp in length  
gap of unknown length  
103740 110294: contig of 6555 bp in length  
gap of unknown length  
110295 116891: contig of 6597 bp in length  
gap of unknown length  
116892 122399: contig of 5508 bp in length  
gap of unknown length  
122400 129404: contig of 7005 bp in length  
gap of unknown length  
129405 136213: contig of 6809 bp in length  
gap of unknown length  
136214 142142: contig of 5929 bp in length  
gap of unknown length  
142143 150401: contig of 8259 bp in length  
gap of unknown length  
150402 158946: contig of 8545 bp in length  
gap of unknown length  
158947 167955: contig of 9009 bp in length  
gap of unknown length  
167956 177918: contig of 9963 bp in length  
gap of unknown length  
177919 183990: contig of 6072 bp in length  
gap of unknown length  
183991 191939: contig of 7949 bp in length  
gap of unknown length  
191940 201055: contig of 9116 bp in length  
gap of unknown length  
201056 210412: contig of 9357 bp in length  
gap of unknown length

\* 210413 219973: contig of 9561 bp in length  
 \* 219974 228547: gap of unknown length  
 \* 228548 252971: contig of 24424 bp in length  
 \* 252972 284497: gap of unknown length  
 \* 284498 316714: gap of unknown length  
 \* 316715 349919: contig of 33205 bp in length

FEATURES  
 source  
 1. 349919  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="19"  
 /clone="CTC-557J18"

BASE COUNT 96739 a 81094 c 78642 g 92256 t 1188 others  
 ORIGIN

Query Match 9.7% Score 69.8; DB 54; Length 349919;  
 Best Local Similarity 57.2%; Pred. No. 0.29;  
 Matches 147; Conservative 0; Mismatches 107; Indels 3; Gaps 1;  
 Qy 50 actaattatttgggttaataataacattatttttaagatat---taattaagaatt 106  
 Db 309365 ACAACAATCTCGTTAAATGAAATGGTTTATAATATATATATATATATAT 309306  
 Qy 107 aaagattttttaaataatgataaaattattattcatgatttttcacattgat 166  
 Db 309305 TAT 309246  
 Qy 167 ttgataataaataattttttttaaattcttcaaaaagtgtgcaaacaccttttagac 226  
 Db 309245 TATAATATAAT 309186  
 Qy 227 atagcttggtctttcacaaagcattcatcatttaacattaaataatttaataac 286  
 Db 309185 TTTATAAT 309126  
 Qy 287 taacagtagaattctt 303  
 Db 309125 ATTTTATAATATATCT 309109

RESULT 6  
 AC026640 164119 bp DNA HTG 13-APR-2000  
 LOCUS  
 DEFINITION Homo sapiens chromosome 11 clone RP11-69K18 map 11, WORKING DRAFT  
 SEQUENCE, 16 unordered pieces.  
 AC026640  
 VERSION AC026640.2 GI:7547222  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavsky,L., Bouckhelter,B., Brown,A., Burkett,G.,  
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,

Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Plisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 13, 2000 this sequence version replaced gi:7284664.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR

Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L8749  
 Center clone name: 69\_K18  
 ----- Summary Statistics

Sequencing vector: M13: M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 150998 bases at least Q40  
 Consensus quality: 158219 bases at least Q30  
 Consensus quality: 160955 bases at least Q20  
 Insert size: 176000; agarose-fp  
 Insert size: 162619; sum-of-contigs  
 Quality coverage: 3.8 in Q20 bases; agarose-fp  
 Quality coverage: 4.1 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 16 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1608: contig of 1608 bp in length  
 \* 1609 1708: gap of 100 bp  
 \* 1709 3184: contig of 1476 bp in length  
 \* 3185 3284: gap of 100 bp  
 \* 3285 5541: contig of 2257 bp in length  
 \* 5542 5641: gap of 100 bp  
 \* 5642 7773: contig of 2132 bp in length  
 \* 7774 7873: gap of 100 bp  
 \* 7874 10690: contig of 2817 bp in length  
 \* 10691 10790: gap of 100 bp  
 \* 10791 13433: contig of 2643 bp in length  
 \* 13434 13533: gap of 100 bp  
 \* 13534 18455: contig of 2922 bp in length  
 \* 16456 16555: gap of 100 bp  
 \* 16556 21573: contig of 5018 bp in length  
 \* 21574 21673: gap of 100 bp  
 \* 21674 25705: contig of 4032 bp in length  
 \* 25706 25805: gap of 100 bp  
 \* 25806 31983: contig of 6178 bp in length  
 \* 31984 32083: gap of 100 bp  
 \* 32084 42802: contig of 10719 bp in length  
 \* 42803 42902: gap of 100 bp  
 \* 42903 54889: contig of 11987 bp in length  
 \* 54890 54989: gap of 100 bp  
 \* 54990 67682: contig of 12693 bp in length  
 \* 67683 67782: gap of 100 bp  
 \* 67783 80005: contig of 12223 bp in length



```

* 27254 27353: gap of 100 bp
* 27354 41791: contig of 14438 bp in length
* 41792 41891: gap of 100 bp
* 41892 61037: contig of 19146 bp in length
* 61038 61137: gap of 100 bp
* 61138 88113: contig of 26976 bp in length
* 88114 88213: gap of 100 bp
* 88214 131619: contig of 43406 bp in length
* 131620 131719: gap of 100 bp
* 131720 179510: contig of 47791 bp in length.
FEATURES
  source
    1. 179510
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="RP11-21P24"
      /clone_lib="RPC1-11 Human Male BAC"
    1. 2601
      /note="assembly_fragment"
    2702. 8297
      /note="assembly_fragment"
    8398. 27253
      /note="assembly_fragment"
    27354. 41791
      /note="assembly_fragment"
      clone_end:SP6
      vector_side:right"
    41892. 61037
      /note="assembly_fragment"
    61138. 88113
      /note="assembly_fragment"
    88214. 131619
      /note="assembly_fragment"
      clone_end:T7
      vector_side:right"
    131720. 179510
      /note="assembly_fragment"
BASE COUNT 59127 a 32374 c 31805 g 55501 t 703 others
ORIGIN

```

```

Query Match          9.6%: Score 69.2; DB 71; Length 179510;
Best Local Similarity 52.8%: Pred. NO. 0.42;
Matches 149; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy 22 ataacatggatctccctaccacataactatttttgggttaataattcaattca 81
Db 3271 ATATAATATATATATATATATATATATATATATATATATATATATATAGAATA 3212
Qy 82 ttttaagatatattaaagatttttaaaaaaagatttttaaaaaaagattttta 141
Db 3211 TATAATATATATATATATATATATATATATATATATATATATATATAGAATA 3152
Qy 142 attcatgatttttcacacattgatttgataataaatatttttttaattttcttaa 201
Db 3151 TTATATATATATATATATATATATATATATATATATATATATATAGAATA 3092
Qy 202 aaatgttgaagacacttattagacatgcttctgttttcacaaagcatcattcat 261
Db 3091 TTATATATATCTAAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 3032
Qy 262 taatacattaaaaatttaataactaacagtagaattctct 303
Db 3031 AATTATATATATATATATATATATATATATATATATATATATATATATATAT 2990

RESULT 8
LOCUS AC023371 48532 bp DNA HTG 14-FEB-2000
DEFINITION Homo sapiens clone RP11-21D18, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC023371
VERSION AC023371.1 GI:6970502
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.

```

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 48532)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

## AUTHORS

Homo sapiens, clone RP11-21D18

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 48532)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barina, N., Beda, F., Boguslavskiy, L., Bouckhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Largocque, K., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Stojanovic, N., Severy, P., Spencer, B., Stange-Thomann, N., Theodore, J., Tirrell, A., Subramanian, A., Talamas, J., Tesfaye, S., Theodores, J., Vo, A., Wilson, B., Travers, M., Triglio, J., Vassiliev, H., Viel, R., Vo, A., Zimmer, A. and Wu, X., Wyman, D., Ye, W. J., Young, G., Zalnoun, J., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

## JOURNAL

Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. &amp; Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submission@genome.wi.mit.edu

----- Project Information

Center project name: L3985

Center clone name: 21\_D\_18

-----

\* NOTE: This record contains 55 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

\* 1 869: contig of 869 bp in length  
 \* 870 1715: contig of 846 bp in length  
 \* 1716 2597: contig of 882 bp in length  
 \* 2598 3468: contig of 871 bp in length  
 \* 3469 4369: contig of 901 bp in length  
 \* 4370 5239: contig of 870 bp in length  
 \* 5240 6103: contig of 864 bp in length  
 \* 6104 6975: contig of 872 bp in length  
 \* 6976 7871: contig of 896 bp in length  
 \* gap of unknown length

```
* 7872 8739: contig of 868 bp in length
* gap of unknown length
* 8740 9613: contig of 874 bp in length
* gap of unknown length
* 9614 10472: contig of 859 bp in length
* gap of unknown length
* 10473 11328: contig of 856 bp in length
* gap of unknown length
* 11329 12188: contig of 860 bp in length
* gap of unknown length
* 12189 13069: contig of 881 bp in length
* gap of unknown length
* 13070 13952: contig of 883 bp in length
* gap of unknown length
* 13953 14829: contig of 877 bp in length
* gap of unknown length
* 14830 15701: contig of 872 bp in length
* gap of unknown length
* 15702 16587: contig of 886 bp in length
* gap of unknown length
* 16588 17461: contig of 874 bp in length
* gap of unknown length
* 17462 18326: contig of 865 bp in length
* gap of unknown length
* 18327 19191: contig of 865 bp in length
* gap of unknown length
* 19192 20081: contig of 890 bp in length
* gap of unknown length
* 20082 20971: contig of 890 bp in length
* gap of unknown length
* 20972 21849: contig of 878 bp in length
* gap of unknown length
* 21850 22713: contig of 864 bp in length
* gap of unknown length
* 22714 23562: contig of 849 bp in length
* gap of unknown length
* 23563 24422: contig of 860 bp in length
* gap of unknown length
* 24423 25286: contig of 864 bp in length
* gap of unknown length
* 25287 26173: contig of 887 bp in length
* gap of unknown length
* 26174 27055: contig of 882 bp in length
* gap of unknown length
* 27056 28036: contig of 981 bp in length
* gap of unknown length
* 28037 28917: contig of 880 bp in length
* gap of unknown length
* 28917 29801: contig of 884 bp in length
* gap of unknown length
* 29801 30670: contig of 870 bp in length
* gap of unknown length
* 30671 31669: contig of 999 bp in length
* gap of unknown length
* 31670 32564: contig of 895 bp in length
* gap of unknown length
* 32565 33435: contig of 871 bp in length
* gap of unknown length
* 33436 34363: contig of 928 bp in length
* gap of unknown length
* 34364 35264: contig of 901 bp in length
* gap of unknown length
* 35265 36131: contig of 867 bp in length
* gap of unknown length
* 36132 37024: contig of 893 bp in length
* gap of unknown length
* 37025 37922: contig of 898 bp in length
* gap of unknown length
* 37923 38829: contig of 907 bp in length
* gap of unknown length
* 38830 39709: contig of 880 bp in length
* gap of unknown length
* 39710 40590: contig of 881 bp in length
```

```
* 40591 41490: contig of 900 bp in length
* gap of unknown length
* 41491 42372: contig of 882 bp in length
* gap of unknown length
* 42373 43253: contig of 881 bp in length
* gap of unknown length
* 43254 44133: contig of 880 bp in length
* gap of unknown length
* 44134 45009: contig of 876 bp in length
* gap of unknown length
* 45010 45881: contig of 872 bp in length
* gap of unknown length
* 45882 46757: contig of 876 bp in length
* gap of unknown length
* 46758 47645: contig of 888 bp in length
* gap of unknown length
* 47646 48532: contig of 887 bp in length.
* Location/Qualifiers
  1..48532
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="RP11-21D18"
    /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 14211 a 9672 c 9208 g 15109 t 332 others
ORIGIN

Query Match 9.6%; Score 69; DB 51; Length 48532;
Best Local Similarity 54.4%; Pred. No. 0.65; Indels 0; Gaps 0;
Matches 135; Conservative 0; Mismatches 113;

Qy 66 taaatattaacattattttaagatatttaataaagaaataaaagattttttaaaaaa 125
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33833 TAAATAATTAAATAATTATTATTAATAATTATATATATTTTAAATAATAT 33892

Qy 126 tgtataaaattatattcatgatgtatttttcatacacattgttttgataataattatt 185
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33893 TATTTATAATATTTTATTATTTTAAATAATAATAATAATATTTTATTAATAATAATAATA 33952

Qy 186 ttttaatttcttaaaaaagtgcagacaccttattagacagtctgtctctttac 245
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33953 TATTTAATATCCCTATCAATATTAATAATAATAATAATAATAATAATAATATTCATNTAA 34012

Qy 246 aaagcattcatcatttaatacacattaaaaataatttaatactaacagtagaattcttg 305
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34013 AAAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 34072

Qy 306 tgagtgggt 313
|||
Db 34073 ATAATNTT 34080
```

```
RESULT 9
DDU87514/c 9
LOCUS DDU87514 975 bp DNA INV 05-APR-1997
DEFINITION Dictyostelium discoideum CAR3 gene, promoter region.
ACCESSION U87514
VERSION U87514.1 GI:1927211
KEYWORDS
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 975)
AUTHORS Kimmel, A.R. and Gollop, R.
TITLE CAR3 promoter of Dictyostelium
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 975)
AUTHORS Kimmel, A.R. and Gollop, R.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1997) LCOB, NIH, 6/BI-22, Bethesda, MD 20892, USA
FEATURES
  Location/Qualifiers
    1..975
  source
```



X04455.1 GI:11640  
 16S ribosomal RNA; 23S ribosomal RNA; 4.5S ribosomal RNA; 5S  
 ribosomal RNA; ATPase; chlorophyll a apoprotein; cytochrome:  
 cytochrome b-559; cytochrome b6; cytochrome f; cytochrome subunit  
 IV; F0-F1-ATPase; Fl-ATPase; initiation factor; photosynthesis;  
 photosystem II; ribosomal protein; ribosomal protein L14; ribosomal  
 protein L16; ribosomal protein L20; ribosomal protein L21;  
 ribosomal protein L22; ribosomal protein L23; ribosomal protein  
 L33; ribosomal protein L11; ribosomal protein S12; ribosomal  
 protein S14; ribosomal protein S15; ribosomal protein S18;  
 ribosomal protein S19; ribosomal RNA; ribulose biphosphate  
 carboxylase; RNA polymerase; transfer RNA; transfer RNA-Ala;  
 transfer RNA-Arg; transfer RNA-Asp; transfer RNA-Cys; transfer  
 RNA-gln; transfer RNA-Glu; transfer RNA-Gly; transfer RNA-His;  
 transfer RNA-Ile; transfer RNA-Leu; transfer RNA-Lys; transfer  
 RNA-Met; transfer RNA-Phe; transfer RNA-Pro; transfer RNA-Ser;  
 transfer RNA-Thr; transfer RNA-Trp; transfer RNA-Tyr; transfer  
 RNA-Val; unidentified reading frame.  
 liverwort.  
 Chloroplast Marchantia polymorpha  
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta:  
 Marchantiophyta: Marchantiales; Marchantiaceae; Marchantia.  
 1 (bases 1 to 121024)  
 Yamano,Y., Ohyama,K. and Komano,T.  
 Nucleotide sequences of chloroplast 5S ribosomal RNA from cell  
 suspension cultures of the liverworts Marchantia polymorpha and  
 Jungermannia subulata  
 Nucleic Acids Res. 12 (11), 4621-4624 (1984)  
 84247325  
 2 (bases 1 to 121024)  
 Umesono,K., Inokuchi,H., Ohyama,K. and Ozeki,H.  
 Nucleotide sequence of Marchantia polymorpha chloroplast DNA: a  
 region possibly encoding three tRNAs and three proteins including  
 1

JOURNAL MEDLINE REFERENCE AUTHORS TITLE	homologue of E. coli ribosomal protein S14 Nucleic Acids Res. 12 (24), 9551-9565 (1984) 85087956 3 (bases 1 to 121024) Fukuzawa, H., Uchida, Y., Yamano, Y., Ohyama, K. and Komano, T. Molecular cloning of promoters, functional in Escherichia coli from chloroplast DNA of a liverwort, Marchantia polymorpha Agric. Biol. Chem. 49, 2725-2731 (1985) 4 (bases 1 to 121024) Yamano, Y., Kohchi, T., Fukuzawa, H., Ohyama, K. and Komano, T. Nucleotide sequences of chloroplast 4.5 S ribosomal RNA from a leafy liverwort, Jungermannia subulata, and Marchantia polymorpha Marchantia polymorpha FEBS Lett. 185, 203-207 (1985) 5 (bases 1 to 121024) Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA Nature 322, 572-574 (1986) 6 (bases 1 to 121024) Fukuzawa, H., Kohchi, T., Shirai, H., Ohyama, K., Umesono, K., Inokuchi, H. and Ozeki, H. Coding sequences for chloroplast ribosomal protein S12 from the liverwort, Marchantia polymorpha, are separated far apart on the different DNA strands FEBS Lett. 198, 11-15 (1986) 7 (bases 1 to 121024) Fukuzawa, H., Yoshida, T., Kohchi, T., Okumura, T., Sawano, Y. and Ohyama, K. Splicing of group II introns in mRNAs coding for cytochrome b6 and subunit IV in liverwort Marchantia polymorpha chloroplast genome: Exon specifying a region coding for two genes with the spacer region FEBS Lett. 220, 61-66 (1987) 8 (bases 1 to 121024) Kohchi, T., Ogura, Y., Umesono, K., Yamada, Y., Komano, T., Ozeki, H. and Ohyama, K. Ordered processing and splicing in a polycistronic transcript in liverwort chloroplasts Curr. Genet. 14 (2), 147-154 (1988) 89028845 9 (bases 1 to 121024) Inokuchi, T., Shirai, H., Fukuzawa, H., Sano, T., Komano, T., Umesono, K., Inokuchi, H., Ozeki, H. and Ohyama, K. Structure and organization of Marchantia polymorpha chloroplast genome. IV. Inverted repeat and small single copy regions J. Mol. Biol. 203 (2), 353-372 (1988) 89068688 10 (bases 1 to 121024) Umesono, K., Inokuchi, H., Shiki, Y., Takeuchi, M., Chang, Z., Fukuzawa, H., Kohchi, T., Shirai, H., Ohyama, K. and Ozeki, H. Structure and organization of Marchantia polymorpha chloroplast genome. II. Gene organization of the large single copy region from rps12 to atpB J. Mol. Biol. 203 (2), 299-331 (1988) 89068686 11 (bases 1 to 121024) Ohyama, K., Fukuzawa, H., Kohchi, T., Sano, T., Sano, S., Shirai, H., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. Structure and organization of Marchantia polymorpha chloroplast genome. I. Cloning and gene identification J. Mol. Biol. 203 (2), 281-298 (1988) 89068685 12 (bases 1 to 121024) Fukuzawa, H., Kohchi, T., Sano, T., Shirai, H., Umesono, K., Inokuchi, H., Ozeki, H. and Ohyama, K. Structure and organization of Marchantia polymorpha chloroplast genome. III. Gene organization of the large single copy region from rbcL to trnI(CAU) J. Mol. Biol. 203 (2), 333-351 (1988)
---	--

## Genes predicted by amino acid sequence homology

ndh1: homologous to mammalian mitochondrial URF1  
 ndh2: URF2 (same as above)  
 ndh3: URF3 (same as above)  
 ndh4: URF4 (same as above)  
 ndh4L: URF4L (same as above)  
 ndh5: URF5 (same as above)  
 URF6: homologous to A.nidulans mitochondrial URF6  
 frxA: homologous to 4Fe-4S type ferredoxin  
 frxB: same as above  
 frxC: homologous to 4Fe-4S protein found in R.capsulata  
 mbpx: homologous to ATP binding subunit of inner membrane

Query Match 9.5%; Score 68.2; DB 8; Length 121024;  
 Best Local Similarity 53.6%; Pred. No. 0.63;  
 Matches 142; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 39 ttaccaatcactaattatttgggtggttaaaatttaattcattatttttaagatattaatt 98  
 DB 32872 TCAAAATAATTTTGGTTATTTAAATAATTTCAATGTTTTTCTTTGTTCAAAAA 32813

QY 99 aagaattaaagattttttaaataaaatgataaaattatatttcaatgatttttcaata 158  
 DB 32812 ATGAAATTAATAATTTATCAAAATTTATATGATTTTCTTAAATAAACATTAGA 32753

QY 159 cattgttttgataaaataatttttttcaatttctttaaataaaatggtgcaagacact 218  
 DB 32752 TAATTGAGATTAAATAAAGTAAATTTTTCATATAAAGATTTAAATTCATTAA 32693

QY 219 tattagacatagcttctgttctgtttacaaaagcattcatttcaatacattaaataaata 278  
 DB 32692 TAATAAAAAATTTCTTTATCTTTTATCTTTTGAAGAAATTTTCAATAATCGTAATAATAATAAAT 32633

QY 279 ttttaatacactaagcagatattcttct 303  
 DB 32632 ATTATTTTAAATAGTTTTTTTCT 32608

RESULT 12  
 CNS01G8P/C 910 bp DNA STS 17-FEB-2000  
 LOCUS Anopheles gambiae STS sp6 end of clone 06E14 of NotreDamel library  
 DEFINITION from strain PEST of Anopheles gambiae (African malaria mosquito),  
 sequence tagged site.

ACCESSION AL142826.1 GI:7000944  
 VERSION STS.  
 KEYWORDS African malaria mosquito.  
 SOURCE Anopheles gambiae  
 ORGANISM Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;  
 Culicoidae; Culicidae; Anophelinae;  
 1 (bases 1 to 910)  
 Genoscope.  
 Direct Submission  
 Submitted (16-FEB-2000) Genoscope - Centre National de Sequençage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)  
 2 (bases 1 to 910)  
 - Web : www.genoscope.cns.fr)

REFERENCE  
 AUTHORS Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissbach,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-FEB-2000) BWM, Institut Pasteur, 25, rue du Dr.  
 Roux, Paris 75015, France  
 COMMENT This clone is from an A. gambiae BAC library provided by F.H.  
 Collins and sequenced by Genoscope in collaboration with the  
 Laboratory of Biochem. and Biol. Molec. of Insects, Institut  
 Pasteur.

## FEATURES

Location/Qualifiers  
 1..910  
 /organism="Anopheles gambiae"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone="06E14"  
 /clone\_lib="NotreDamel"  
 /note="end : sp6"  
 BASE COUNT 376 a 83 c 91 g 294 t 66 others  
 ORIGIN

Query Match 9.4%; Score 67.8; DB 13; Length 910;  
 Best Local Similarity 48.2%; Pred. No. 2.9;  
 Matches 123; Conservative 15; Mismatches 117; Indels 0; Gaps 0;

QY 52 taattatttgggtggttaaaatttcaatttattttaaagattttaaagattttaaag 111  
 DB 896 TAGATATAATTTTAAATTAATTAATAAATAATGATATTAATAATTAATAAATTAATAA 837

QY 112 attttttaaataaaatgataaaatttatttcatgatttcttcaatatttgaatttga 171  
 DB 836 TTATTTTAAWRNATATTTTATATAANANWTGTTWARTAAATTTNTWAAATTTTATTAMNA 777

QY 172 taataaataattttttttaaatttctttaaataaattggtgcaagacactatttagacatag 231  
 DB 776 ATTNTATTTTNTTNTTATTTTATTTATATATATTTTWTWAAATTTTATTTTAT 717

QY 232 ctgttctgtttacaaaagcattcatttcaatacattaaataatttataatacaca 291  
 DB 716 WTATAAGTAAATTAATTTTATATAAATTTATATAAATTTTWTATATAATTAATA 657

QY 292 gtagaattcttctgt 306  
 DB 656 AAATTAATAATTTT 642

RESULT 13  
 CNS01DX9 99263 bp DNA HTG 11-FEB-2000  
 LOCUS Homo sapiens chromosome 14 clone R-993B13, \*\*\* SEQUENCING IN  
 DEFINITION PROGRESS \*\*\*, 2 ordered pieces.

ACCESSION AL139177.1 GI:6983341  
 VERSION HTG.  
 KEYWORDS HTGS\_PHASE2; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Genoscope.  
 Direct Submission  
 Submitted (11-FEB-2000) Genoscope - Centre National de Sequençage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)  
 - Web : www.genoscope.cns.fr)

COMMENT On Feb 16, 2000 this sequence version replaced gi:6977814.  
 IMPORTANT: This sequence is unfinished and does not necessarily  
 represent the correct sequence. Work on the sequence is in progress  
 and the release of this data is based on the understanding that the  
 sequence may change as work continue. The sequence may be  
 contaminated with foreign sequence from E.coli, yeast, vector,  
 phage, etc.  
 Contig order : 3 4, 1000 N's separate segments  
 Contig 3 : length 4694 bp.  
 Contig 4 : length 93569 bp.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and

```

* the accession number will be preserved.
* 1 4694: contig of 4694 bp in length
* 4695 5694: gap of 1000 bp
* 5695 99263: contig of 93569 bp in length.
Location/Qualifiers
1. 99263
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone_lib="RPCI-11"
/clone="R-993B13"
3187. 3395
/standard_name="Z16496 (RH53488 STS20513)"
/note="matching EMBL:Z16496; dbSTS:STS20513; Identified
using the e-PCR software (G. Schuler)"
79452. 79632
/standard_name="G03454 (RH53911 STS2670)"
/note="matching EMBL:G03454; dbSTS:STS2670; Identified
using the e-PCR software (G. Schuler)"
94533. 94651
/standard_name="N57675 (RH65918 STS45848)"
/note="matching EMBL:N57675; dbSTS:STS45848; Identified
using the e-PCR software (G. Schuler)"
BASE COUNT 31240 a 19949 c 18494 g 28472 t 1108 others
ORIGIN

Query Match
Best Local Similarity 45.6%; Pred. No. 1;
Matches 187; Conservative 0; Mismatches 216; Indels 7; Gaps 1;

QY 3 atcaactcacatccaaacataacatggatctctcctaccatcatcattattttg 62
|| || | || || || || || || || || || || || || || || || ||
Db 4202 ATAAAAATATTTTAAAAATNANCTCATTTCTNTATTTNATTAATAAATAT 4261

QY 63 gggttaataattcatctatttttaagatatttaataagaaatttaattttaa 122
|| || | || || || || || || || || || || || || || || || ||
Db 4262 TAAAAATTTTAAAAATATTTTAAATANAANAANAANAANAANAANAANA 4321

QY 123 aaatgtataaaattattattcatgatttttcacatttgatttgataataata 182
|| || | || || || || || || || || || || || || || || || ||
Db 4322 AANTNTNAATAAAAAAATACCATTTAATAATAAATTTTAAATANAANAANA 4381

QY 183 ttttttttaattcttaaaaaatgttgcagacacttattagacatgcttctgt 242
|| || | || || || || || || || || || || || || || || || ||
Db 4382 TTTATTATTATTNTTNAACA-----AATATAATTAATAAAAAAATAAAT 4434

QY 243 tacaagaagcatcattatttaataacatttaataatttaataacagtaga 302
|| || | || || || || || || || || || || || || || || || ||
Db 4435 TTAANAANAANAANAANAATAAANAACAANAANAANAANAANAANAANA 4494

QY 303 ttgtgagtgtgtggagtaggcaacctggcattgaaacgagagaagagagc 362
|| || | || || || || || || || || || || || || || || || ||
Db 4495 TTATATATATTTTATATAAATATANANNNTANANNNTANANNNTANANN 4554

QY 363 agaagacaaataaaagtagtgcacaaacaaatcacaagaggagcaaa 412
|| || | || || || || || || || || || || || || || || || ||
Db 4555 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 4604

RESULT 14
HS1108D11 152209 bp DNA PRI. 17-MAR-2000
LOCUS
DEFINITION
Human DNA sequence from clone RP5-1108D11 on chromosome
20q12-13.11. Contains part of the gene for a novel protein similar
to C. elegans T2C1.7, part of the gene for a novel HMG (high
mobility group) box protein similar to KIAA0737, KIAA0808 and TNRC9
(CAGF9), ESTs, STSs, GSs and two putative CpG islands, complete
sequence.
ACCESSION AL034419.19 GI:6933861
VERSION AL034419
KEYWORDS HTG: CAGF9; CpG island; HMG; KIAA0737; KIAA0808; TNRC9.
SOURCE human.

```

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 152209)

Clark.G.

Direct Submission

Submitted (13-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Feb 7, 2000 this sequence version replaced gi:6911915.

During sequence assembly data is compared from overlapping clones.

where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence has been finished according to sequence map criteria

as follows. An attempt is made to resolve all sequencing problems,

such as compressions and repeats, but not necessarily within known

annotated human repeat sequence elements (e.g. Alu). Where the

sequence is ambiguous, there is an annotation using the 'unsure'

feature key.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C\_elegans/wormpep/

This sequence was generated from part of bacterial clone contigs of human

chromosome 20, constructed by the Sanger Centre Chromosome 20

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr20

RP5-1108D11 is from the library RPCI-5 constructed at the Roswell

Park Cancer Institute by the group of Pieter de Jong. For further

details see http://bacpac.med.buffalo.edu/

VECTOR: pcypac2

This sequence is the entire insert of clone RP5-1108D11. The true

left end of clone RP3-49503 is at 114928 in this sequence. The true

right end of clone RP5-1183121 is at 70703 in this sequence.

## FEATURES

.source

1. 152209

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="20"

/map="q12-q13.11"

/clone="RP5-1108D11"

/clone\_lib="RPCI-5"

310. 562

/note="L2 repeat: matches 2500. 2748 of consensus"

565. 657

repeat\_region

/note="LIMA2 repeat: matches 6219. 6304 of consensus"

678. 731

repeat\_region

/note="MLT2A repeat: matches 400. 453 of consensus"

732. 755

repeat\_region

/note="12 copies 2 mer gt 100% conserved"

780. 1183

repeat\_region

/note="MLT2A repeat: matches 1. 413 of consensus"

1189. 1271

repeat\_region

/note="L2 repeat: matches 2370. 2477 of consensus"

1634. 1803

repeat\_region

/note="L2 repeat: matches 1533. 1719 of consensus"

1901. 2117

repeat\_region

/note="MIR repeat: matches 2. 218 of consensus"

2467. 2731

repeat\_region

/note="AluJb repeat: matches 11. 283 of consensus"

3015. 3966

misc\_feature

/note="CpG island"

/evidence=not\_experimental

3136. 45038

gene

/gene="dJ1108D11.1"

Join(&lt;3136. 3925,44920. &gt;45038)

/gene="dJ1108D11.1"

/note="match: proteins: Tr:Q22667"

/codon\_start=3

CDS

```
/evidence-not_experimental
/product="d1108D11.1 (novel protein similar to C. elegans
T22C1.7 (Tr:Q22667) (part of protein d118121.2))"
/protein_id="CAB81659.1"
/db_xref="GI:7263911"
/translation="NYQQTNGMHRGVSQVPGYGVVVRSLR7SLSLRSEHS
NGTVAPDPSASPASDPALPSPAIPIRGGFALSLANAAARAPKGGIFQRCALLGK
LRRASGTVSQRVSRLKSLSDSSGASDAASTASLGEAAEGADEAPFEADIDATT
TETYGWKNDRSGFVSERSGLRYEGEMLDNRHGYGCTTLPDGHRECKYRHHV
LVKDTKRMALQKSNKVRQKVEHSVEGAQRAAAIARQAEIAASTSHAKAKAEAEQ
AALAAQESNTARTLARELPDYCP"
repeat_region 3588..3794
/feature="3 copies 69 mer 72% conserved"
repeat_region 4958..5135
/feature="MIR repeat: matches 69..252 of consensus"
repeat_region 5428..5566
/feature="MIR repeat: matches 113..254 of consensus"
repeat_region 5733..5845
/feature="MIR repeat: matches 25..136 of consensus"
repeat_region 5846..5921
/feature="MIR repeat: matches 63..139 of consensus"
repeat_region 5977..6392
/feature="L1R16A repeat: matches 16..450 of consensus"
misc_feature complement(5998..6470)
/feature="match: GSS: Em:AQ812840"
repeat_region 6469..6548
/feature="MIR repeat: matches 63..142 of consensus"
repeat_region 6709..7083
/feature="L1R1A1 repeat: matches 3..394 of consensus"
repeat_region 7106..7166
/feature="MER94 repeat: matches 4..64 of consensus"
repeat_region 7179..7493
/feature="L1MD3 repeat: matches 7431..7739 of consensus"
repeat_region 7494..7789
/feature="AlusX repeat: matches 1..296 of consensus"
repeat_region 7790..8083
/feature="L1MD3 repeat: matches 7034..7431 of consensus"
repeat_region 8189..8388
/feature="MIR repeat: matches 8..233 of consensus"
repeat_region 9046..9363
/feature="MER33 repeat: matches 1..324 of consensus"
repeat_region 10136..10308
/feature="MIR repeat: matches 47..226 of consensus"
repeat_region 10950..11011
/feature="MIR repeat: matches 192..250 of consensus"
repeat_region 11012..11147
/feature="BC200 repeat: matches 3..136 of consensus"
repeat_region 11153..11335
/feature="MIR repeat: matches 11..201 of consensus"
repeat_region 11353..11800
/feature="L2 repeat: matches 2234..2705 of consensus"
repeat_region 12005..12308
/feature="AlusX repeat: matches 1..304 of consensus"
repeat_region 12374..12467
/feature="L1MC5 repeat: matches 7833..7925 of consensus"
repeat_region 12487..12799
/feature="AluJB repeat: matches 1..309 of consensus"
repeat_region 12830..13292
/feature="L1M4 repeat: matches 4351..4825 of consensus"
repeat_region 13358..13529
/feature="L1MC5 repeat: matches 7249..7419 of consensus"
repeat_region 13552..13845
/feature="AlusX repeat: matches 2..309 of consensus"
repeat_region 14067..14098
/feature="16 copies 2 mer ac 84% conserved"
repeat_region 14219..14348
/feature="L2 repeat: matches 1995..2139 of consensus"
repeat_region 14432..14761
/feature="AluY repeat: matches 1..310 of consensus"
repeat_region 15256..15341
/feature="MIR repeat: matches 156..251 of consensus"
repeat_region 15342..15662
/feature="AlusX repeat: matches 1..311 of consensus"
```

```
repeat_region 15663..15739
/feature="MIR repeat: matches 40..156 of consensus"
repeat_region 15903..16262
/feature="L1ME3A repeat: matches 5678..6018 of consensus"
misc_feature complement(16083..16468)
/feature="match: GSS: Em:AQ262524"
repeat_region 17211..17513
/feature="AluB8 repeat: matches 1..310 of consensus"
misc_feature complement(17463..17840)
/feature="match: GSS: Em:AQ776838"
repeat_region 17657..17742
/feature="MIR repeat: matches 59..144 of consensus"
repeat_region 17919..18220
/feature="AlusX repeat: matches 1..303 of consensus"
repeat_region 18485..18670
/feature="MIR repeat: matches 46..228 of consensus"
repeat_region 18705..18811
/feature="MIR repeat: matches 123..234 of consensus"
misc_feature complement(18708..19271)
/feature="match: GSS: Em:AQ760123"
misc_feature complement(18812..19279)
/feature="match: GSS: Em:AQ777683"
misc_feature complement(18825..19251)
/feature="match: GSS: Em:B90590"
repeat_region 18964..19192
/feature="MIR repeat: matches 6..247 of consensus"
repeat_region 19623..19802
/feature="L2 repeat: matches 1997..2185 of consensus"
repeat_region 19812..19999
/feature="L1R16C repeat: matches 202..384 of consensus"
repeat_region 20267..20585
/feature="L2 repeat: matches 2206..2545 of consensus"
repeat_region 20733..21144
/feature="L2 repeat: matches 2264..2706 of consensus"
repeat_region 21176..21340
/feature="MER33 repeat: matches 10..205 of consensus"
repeat_region 21735..22030
/feature="AluJ repeat: matches 1..296 of consensus"
repeat_region 22059..22239
/feature="MLT1J repeat: matches 250..439 of consensus"
repeat_region 22345..22416
/feature="MLT1J repeat: matches 1..72 of consensus"
repeat_region 22522..22556
/feature="MLT1J repeat: matches 1..72 of consensus"

Query Match 9.3% Score 66.8; DB 11; Length 152209;
Best Local Similarity 52.3%; Pred. No. 0.88;
Matches 146; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 27 atggatatctccctaccataactactaatattttgggttaaatataatcattattttt 86
Db 29312 ATATATATTTAATAAATAATATATATATATATATATATATATATATATATAT 29371

QY 87 aagatattaataagaataataagatttttaaaaaataataataataataatattca 146
Db 29372 TAAATATTTATATTTATTAATAAATAATTTATTTATATATATATATATATAT 29431

QY 147 tgattttcacaattgattttgataaataataataataataataataataataataa 206
Db 29432 TATATTTAATAAATAATATATATTTAATAAATAATATATATATATATATATAT 29491

QY 207 ttcaagacactatttagacatagctgtctgtttacaagaagcattcatcattaa 266
Db 29492 ATATTTATATATAAATAATATATTTATATATATATATATATATATATATATAT 29551

QY 267 cattaataataatttaataactaacagtagaactctttt 304
Db 29552 TATATATATATTTATATATATATATATATATATATATATATATATATATAT 29589
```

RESULT 15

MTSCAJ23

LOCUS

DEFINITION

MTSCAJ23 1867 bp DNA circular PLN 11-MAR-1998  
Saccharomyces cerevisiae mitochondrial rRNA-Tyr, tRNA-Asn, &

tRNA-Met genes.

ACCESSION AJ223323

VERSION AJ223323.1 GI:2879886

KEYWORDS transfer RNA-Asn; transfer RNA-Met; transfer RNA-Tyr; tRNA-Asn gene; tRNA-Met gene; tRNA-Tyr gene.

SOURCE baker's yeast.

ORGANISM Mitochondrion Saccharomyces cerevisiae

REFERENCE 1 (bases 1 to 1867)

AUTHORS Francisci, S.

TITLE Direct Submission

JOURNAL Submitted (05-FEB-1998) Francisci S., Dept. of Cell and Developmental Biology, University of Rome I, P.ale A. Moro 5 Rome, 00185, ITALY

REFERENCE 2 (bases 1 to 1867)

AUTHORS Pistilli, D.

JOURNAL Thesis (1997) Dept. of Cell & Developmental Biology, University of Rome I, Rome, ITALY

FEATURES

source Location/Qualifiers

1..1867

/organism="Saccharomyces cerevisiae"

/organelle="mitochondrion"

/strain="D273-10B"

/db\_xref="taxon:4932"

/haplotype="ade5"

/sex="alpha"

1..84

/gene="tRNA-Tyr"

/note="codon recognized: UAC"

/product="tRNA-Tyr"

/anticodon="(pos:37..39,aa:Tyr)"

1..84

/gene="tRNA-Tyr"

613..684

/gene="tRNA-Asn"

/note="codon recognized: AGC"

/product="tRNA-Asn"

/anticodon="(pos:34..36,aa:Asn)"

613..684

/gene="tRNA-Asn"

1794..1867

/gene="tRNA-Met"

/note="codon recognized: CAG"

/product="tRNA-Met"

/anticodon="(pos:35..37,aa:Met)"

1794..1867

/gene="tRNA-Met"

BASE COUNT 815 a 89 c 104 g 859 t

ORIGIN

Query Match 9.3%; Score 66.6; DB 7; Length 1867;

Best Local Similarity 53.7%; Pred. NO. 3.3;

Matches 138; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 27 atggatattcctccaccacatacctaattattttgggttaataattacattattttt 86

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 1438 ATATTTTATTTTAAATTAACAATATATTATTTTAAATTAATTAACCTTTA 1497

QY 87 agatattataaagaattaaaagatttttataaaaaatgtataaaattattattca 146

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 1498 TATATATATATATATATATATTTTAAATTAATTAATTAATATATCTTTTTTTTATAATAA 1557

QY 147 tgatttttcacatttgatttgataataatatttttttaattttcttaaaaaatg 206

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 1558 TCATAATATATATATATTTTATATTAAGATTAAATTTTATATATTATTTT 1617

QY 207 ttgcagacacttattagacatgtcttctgtttacaaaagcattcatcatttaata 266

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 1618 TTTTATTAATAATTTATTTATTAATTTATTTATTTATTTATTTATTTATTAAGAATA 1677

QY 267 cattaaaaaatatttaa 283

Db 1678 TATTTATTAATTTTA 1694

Search completed: September 2, 2000, 03:05:39

Job time: 12803 sec